MACHINE LEARNING AND OPTIMIZATION IN NEUROSCIENCE

Concha Bielza, Pedro Larrañaga

Computational Intelligence Group
Artificial Intelligence Department
Technical University of Madrid, Spain

CAEPIA’11
Tenerife, 7 de noviembre de 2011
Outline

1. Introduction

2. Standard machine learning methods
   - Clustering for neuron classification
   - Clustering of spines
   - Knowledge discovery in Alzheimer’s disease
   - Supervised classification of neurons
   - Supervised classification of dementia development in Parkinson’s
   - Supervised classification for outcome prediction of epilepsy surgery

3. Advanced statistical and machine learning methods
   - EM based subspace clustering for discovering new types of neurons
   - Bayesian classifiers for probabilistic class neuron
   - Multi-dimensional classification for EQ-5D health states in Parkinson
   - Computer simulation of dendritic morphology with Bayesian networks
   - Spatial point processes for distribution of synapses
   - Local regularized regression to decode cognitive states in fMRI experiments

4. Optimization
   - Categorizing Parkinson disease based on non-motor symptoms
   - EDAs for parameter optimization in compartmental model
   - Multi-objective GAs for brain networks
   - Classifying MEG data with lasso logistic regression and EDAs

5. Conclusions
Outline

1. Introduction
   - Standard machine learning methods
     - Clustering for neuron classification
     - Clustering of spines
     - Knowledge discovery in Alzheimer’s disease
     - Supervised classification of neurons
     - Supervised classification of dementia development in Parkinson’s
     - Supervised classification for outcome prediction of epilepsy surgery

3. Advanced statistical and machine learning methods
   - EM based subspace clustering for discovering new types of neurons
   - Bayesian classifiers for probabilistic class neuron
   - Multi-dimensional classification for EQ-5D health states in Parkinson
   - Computer simulation of dendritic morphology with Bayesian networks
   - Spatial point processes for distribution of synapses
   - Local regularized regression to decode cognitive states in fMRI experiments

4. Optimization
   - Categorizing Parkinson disease based on non-motor symptoms
   - EDAs for parameter optimization in compartmental model
   - Multi-objective GAs for brain networks
   - Classifying MEG data with lasso logistic regression and EDAs

5. Conclusions
Blue Brain Project

- Founded in 2005 by the Brain and Mind Institute of the École Polytechnique in Lausanne (Switzerland), directed by Henry Markram
- An attempt to reverse-engineer the mammalian brain, in order to understand brain function and dysfunction through detailed simulations
- One of the 14 grand challenges for Engineering in the 21st century
  [http://www.engineeringchallenges.org](http://www.engineeringchallenges.org)
- Interest in neocortical column, the smallest functional unit of the neocortex
- ~2 million columns in humans, 2mm tall
Column simulation of BBP
At the end of 2008, Universidad Politécnica de Madrid (UPM) and Instituto Cajal (IC) from the Spanish Research Council formed the subproject Cajal Blue Brain, until 2018.

- UPM, including Madrid Supercomputing and Visualization Center: data analysis, optimization, image analysis and visualization
- IC: morphology and function of neuronal cells
The human brain

- The most **complex** organ of our central nervous system and the most complex biological structure known
- **Weight = 1.5kg, width = 140 mm, length = 167mm, height = 93mm**
- **Consumes 20% of total body oxygen**
- **Stops growing at age 18**
The human brain

Brain lobes and layers

- Left-right hemispheres covered by a thin layer of gray matter known as the cerebral cortex, the most recently evolved region of the vertebrate brain, divided into 4 areas
- Some areas, such as the cortex and cerebellum, consist of layers, folded to fit within the available space
- In mammals, neocortex: complex 6-layered structure
- Outermost=layer 1, innermost=layer 6
The human brain

Functions of cortex (as currently understood)

- **Occipital** lobe: visual information
- **Temporal** lobe: auditory signals, processing language, meaning of words
- **Parietal** lobe: touch, taste, pressure, pain, temperature
- **Frontal** lobe: motor activity, integration of muscle activity, speech, thought
- Remaining parts associated with higher thought processes, planning, memory, personality and other human activities
The human brain

Brain at microscopic level

- Composed of **neurons**, blood vessels, glial cells (supporting cells of the CNS)
- Neuron is the basic structural and functional unit of the nervous system —**neuron doctrine**— (S. Ramón y Cajal, late 19th century)
- Just **4 microns** thick → could fit 30,000 neurons on the head of a pin
- ~**100,000 million neurons** (more than known stars in the universe)
3 parts of a neuron

1. **Dendrites** receive info from another cell and transmit the message to soma
2. **Cell body** (or **soma**) contains the nucleus, mitochondria and other organelles typical of eukaryotic cells
3. **Axon** conducts messages away to the next neuron via a specialized structure—**synapse**—
   - Axons fill most of the space in the brain → >150,000 km in the human brain!!
   - Each neuron connected to 1,000 neighboring neurons
   - 10,000 synaptic connections each
Observing the neurons

Optical (or light) microscope. Stain the tissue

Magnify image up to 2000 times

Golgi’s method (1873) stains some cells at random avoiding much overlap
Observing the neurons

Modern electron microscope

Magnify image up to 2 million times

3D from multiple 2D images
“Visualizing” mental activities from brain images

Electrical activity directly or indirectly

- Electroencephalography (EEG)
- Positron Emission Tomography (PET)
- Magnetic Res. Imaging (MRI)
- Functional NIR Spectroscopy (fNIRS)
- Single Photon Emission Computed Tomography (SPECT)
- Functional MRI (fMRI)
Outline

1. Introduction

2. Standard machine learning methods
   - Clustering for neuron classification
   - Clustering of spines
   - Knowledge discovery in Alzheimer’s disease
   - Supervised classification of neurons
   - Supervised classification of dementia development in Parkinson’s
   - Supervised classification for outcome prediction of epilepsy surgery

3. Advanced statistical and machine learning methods
   - EM based subspace clustering for discovering new types of neurons
   - Bayesian classifiers for probabilistic class neuron
   - Multi-dimensional classification for EQ-5D health states in Parkinson
   - Computer simulation of dendritic morphology with Bayesian networks
   - Spatial point processes for distribution of synapses
   - Local regularized regression to decode cognitive states in fMRI experiments

4. Optimization
   - Categorizing Parkinson disease based on non-motor symptoms
   - EDAs for parameter optimization in compartmental model
   - Multi-objective GAs for brain networks
   - Classifying MEG data with lasso logistic regression and EDAs

5. Conclusions
Outline

1 Introduction

2 Standard machine learning methods
   - Clustering for neuron classification
   - Clustering of spines
   - Knowledge discovery in Alzheimer’s disease
   - Supervised classification of neurons
   - Supervised classification of dementia development in Parkinson’s
   - Supervised classification for outcome prediction of epilepsy surgery

3 Advanced statistical and machine learning methods
   - EM based subspace clustering for discovering new types of neurons
   - Bayesian classifiers for probabilistic class neuron
   - Multi-dimensional classification for EQ-5D health states in Parkinson
   - Computer simulation of dendritic morphology with Bayesian networks
   - Spatial point processes for distribution of synapses
   - Local regularized regression to decode cognitive states in fMRI experiments

4 Optimization
   - Categorizing Parkinson disease based on non-motor symptoms
   - EDAs for parameter optimization in compartmental model
   - Multi-objective GAs for brain networks
   - Classifying MEG data with lasso logistic regression and EDAs

5 Conclusions
Clustering to classify interneuron types

- Still unclear **how many subtypes** of cortical neurons exist
- Crucial for understanding how cortical circuits are connected and do their job
- In the past, cell type classification was **qualitative** and subjective leading to inconsistent classes of neurons
- Several recent attempts to classify neurons **quantitatively** from **multimodal information** (physiological, morphological, molecular)
- GABA-ergic neocortical **interneurons** a particularly difficult case, due to their greater physiological, morphological, molecular diversity

Main interneuron types of layers II-IV
Affinity propagation to classify interneuron types

Affinity propagation for classification of neuronal cell types: A case study on four interneuron cell types

*in preparation*
Affinity propagation

Clustering by passing messages between data points [Frey and Dueck, 2007]

- \( k \)-means quite sensitive to the initial selection of exemplars (not always solved by repeating with different initializations). **Chaining effect** of hierarchical
- AP considers simultaneously all data points as **exemplar candidates** and gradually identifies clusters
- **Lower error** than other methods and **faster**

**Input:**
- Matrix of similarities \( s(i, k) \) (how well point \( k \) is suited to be the exemplar for point \( i \));
- Preferences \( s(k, k) \), larger for points more likely to be exemplars
- 2 types of messages between pairs of nodes, point \( i \) and candidate exemplar \( k \)

Each message reflects the current **affinity** that one point \( i \) has for choosing another point \( k \) as its exemplar

- **Responsibility:** \( r(i, k) \leftarrow s(i, k) - \max_{k' \neq k} \{ a(i, k') + s(i, k') \} \) from \( i \) to \( k \): how strongly \( i \) favors \( k \) over other candidates
- **Availability:** \( a(i, k) \leftarrow \min \left\{ 0, r(k, k) + \sum_{i' \neq i, k} \max\{0, r(i', k)\} \right\} \) (initialized as 0) from \( k \) to \( i \): how much \( k \) is available as an exemplar for \( i \)
- **Self-availability** is updated differently: \( a(k, k) \leftarrow \sum_{i' \neq k} \max\{0, r(i', k)\} \)
Affinity propagation

Clustering by passing messages between data points [Frey and Dueck, 2007]

Closer to red indicates more evidence to be an exemplar; darker arrow indicates more strength of the message

Until a stopping criterion is met, update responsibilities given availabilities and availabilities given responsibilities, and for each $i$ choose $k$ (as exemplar) that maximizes $[a(i, k) + r(i, k)]$

Stop after a fixed number of iterations, after low changes in the messages, or after local decisions stay constant during some iterations
AP to classify interneuron types

- 50 interneurons from mice viewed with light microscope, properly stained, then reconstructed using Neurolucida software
- 67 morphological (with the Neurolucida Explorer program) variables and...
- 20 physiological (electrical activity of neurons with current injections) variables
### Neurolucida (morphological variables)

<table>
<thead>
<tr>
<th><strong>DENDRITIC VARIABLES</strong></th>
<th><strong>AXONAL VARIABLES</strong></th>
<th><strong>SOMATIC VARIABLES</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>number of dendrites</td>
<td>axonal node total</td>
<td>somatic perimeter</td>
</tr>
<tr>
<td>dendritic node total</td>
<td>total axonal length</td>
<td>somatic area</td>
</tr>
<tr>
<td>total dendritic length</td>
<td>ratio of axonal length to surface</td>
<td>somatic aspect ratio</td>
</tr>
<tr>
<td>average length of dendrites</td>
<td>highest order axon segment</td>
<td>somatic compactness</td>
</tr>
<tr>
<td>total surface area of dendrites</td>
<td>axonal torsion ratio</td>
<td>somatic form factor</td>
</tr>
<tr>
<td>ratio of dendritic length to surface</td>
<td>axonal planar angle ave</td>
<td>somatic roundness</td>
</tr>
<tr>
<td>dendritic torsion ratio</td>
<td>axonal planar angle stdv</td>
<td>relative distance to pia</td>
</tr>
<tr>
<td>dendritic planar angle ave</td>
<td>axonal local angle ave</td>
<td></td>
</tr>
<tr>
<td>dendritic planar angle stdv</td>
<td>axonal local angle stdv</td>
<td></td>
</tr>
<tr>
<td>dendritic local angle ave</td>
<td>axonal spline angle ave</td>
<td></td>
</tr>
<tr>
<td>dendritic local angle stdv</td>
<td>axonal spline angle stdv</td>
<td></td>
</tr>
<tr>
<td>dendritic spline angle ave</td>
<td>ave tortuosity of axonal segments</td>
<td></td>
</tr>
<tr>
<td>dendritic spline angle stdv</td>
<td>stdv of tortuosity of axonal segments</td>
<td></td>
</tr>
<tr>
<td>ave tortuosity of dendritic seg</td>
<td>axonal segment length ave</td>
<td></td>
</tr>
<tr>
<td>stdv of tortuosity of dendritic seg</td>
<td>axonal segment length stdv</td>
<td></td>
</tr>
<tr>
<td>dendritic segment length ave</td>
<td>ave tortuosity of axonal nodes</td>
<td></td>
</tr>
<tr>
<td>dendritic segment length stdv</td>
<td>stdv tortuosity of axonal nodes</td>
<td></td>
</tr>
<tr>
<td>ave tortuosity of dendritic nodes</td>
<td>number axonal sholl sections</td>
<td></td>
</tr>
<tr>
<td>stdv of tortuosity of dendritic nodes</td>
<td>axonal sholl length at 100 µm</td>
<td></td>
</tr>
<tr>
<td>number of dendritic sholl sections</td>
<td>axonal sholl length at 200 µm</td>
<td></td>
</tr>
<tr>
<td>dendritic sholl length at 50 µm</td>
<td>axonal sholl length at 300 µm</td>
<td></td>
</tr>
<tr>
<td>dendritic sholl length at 100 µm</td>
<td>axonal length density2</td>
<td></td>
</tr>
<tr>
<td>dendritic sholl length at 150 µm</td>
<td>axonal node density2</td>
<td></td>
</tr>
<tr>
<td>convex hull dendrite area</td>
<td>convex hull axon area</td>
<td></td>
</tr>
<tr>
<td>convex hull dendrite perimeter</td>
<td>convex hull axon perimeter</td>
<td></td>
</tr>
<tr>
<td>convex hull dendrite volume</td>
<td>convex hull axon volume</td>
<td></td>
</tr>
<tr>
<td>convex hull dendrite surface area</td>
<td>convex hull axon surface area</td>
<td></td>
</tr>
<tr>
<td>highest order dendritic segment</td>
<td>k-dim axon (fractal analysis)</td>
<td></td>
</tr>
<tr>
<td>k-dim dendrites (fractal analysis)</td>
<td>total surface area of axon</td>
<td></td>
</tr>
</tbody>
</table>
*AP to classify interneuron types*

**Results of AP**

- **Similarities**: defined from different distances (Euclidean, sEuclidean, cosine, Pearson correlation, Spearman correlation...)
- **Equal preferences**: the median of all similarity values
- **Both issues** affect the number of clusters and their arrangement
- 4 **known** subtypes: parvalbumin-positive (PV) basket cells, PV+ chandelier cells (CC), somatostatin-positive (SOM-Mt) Martinotti cells, and SOM+ non-Martinotti group 2/3 cells (SOM-2-3)
- For Spearman measure (the most consistent):

<table>
<thead>
<tr>
<th></th>
<th>Combined (8 clusters)</th>
<th>Morphological (9 clusters)</th>
<th>Physiological (6 clusters)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>0.82</td>
<td>0.82</td>
<td>0.76</td>
</tr>
<tr>
<td>M2</td>
<td>0.79</td>
<td>0.78</td>
<td>0.733</td>
</tr>
<tr>
<td>M3</td>
<td>0.82</td>
<td>0.82</td>
<td>0.76</td>
</tr>
</tbody>
</table>

M1: neuron is assigned to the class of its exemplar (Acc = correctly classified neurons divided by the number of neurons)
M2: neuron is assigned to the class of its exemplar (Acc excluding the exemplars)
M3: neuron is assigned to the majority class in its cluster (Acc as in M1)
AP to classify interneuron types

The 6 clusters found by AP using physiological variables

- PV, CC, SOM-Mt, SOM-2-3

- Misclassified cells were most often of the same general class of PV+ or SOM+.
  In fact, Acc with 2 general classes (PV+ and SOM+) was higher than 0.97

- AP outperformed hierarchical clustering using Ward’s method
Outline

1. Introduction

2. Standard machine learning methods
   - Clustering for neuron classification
   - Clustering of spines
   - Knowledge discovery in Alzheimer’s disease
   - Supervised classification of neurons
   - Supervised classification of dementia development in Parkinson’s
   - Supervised classification for outcome prediction of epilepsy surgery

3. Advanced statistical and machine learning methods
   - EM based subspace clustering for discovering new types of neurons
   - Bayesian classifiers for probabilistic class neuron
   - Multi-dimensional classification for EQ-5D health states in Parkinson
   - Computer simulation of dendritic morphology with Bayesian networks
   - Spatial point processes for distribution of synapses
   - Local regularized regression to decode cognitive states in fMRI experiments

4. Optimization
   - Categorizing Parkinson disease based on non-motor symptoms
   - EDAs for parameter optimization in compartmental model
   - Multi-objective GAs for brain networks
   - Classifying MEG data with lasso logistic regression and EDAs

5. Conclusions
Clustering of spines

- Spines found on the dendrites of most principal neurons. A human apical dendrite could have 6000 spines
- Where synapses typically formed (discovered by Ramón y Cajal)
- Spine plasticity implicated in learning and memory
- Many different shapes. Strong synaptic contacts ↔ large spine head; age- and disease-related declines in cognitive ability ↔ lower spine density

Basic classes: stubby (1), thin (2), mushroom (9,10,11), ramified (15) [Peters and Kaiserman-Abramof, 1970], based on optical microscopy

But most are atypical or intermediate types (3-8, 12-14) [Arellano et al., 2007], based on electron microscopy (a continuum with no clear sub-grouping??)

Cluster spines with their morphological parameters to improve the classification proposed by Peters and Kaiserman-Abramof
Clustering of spines

Clustering of spines in pyramidal neurons according to their morphology
in preparation
Clustering of spines

7543 spines from 2 healthy human subjects (40 and 85 y.o.) from the same cortical area, based on confocal laser microscopy images.

Manually segmented, a semiautomatic procedure that took almost 3 years of work, to correctly select and count the objects of interest, distinguishing from noise.

- Select the image intensity threshold for each spine or pieces of spine, mark the point of insertion in the dendritic body and mark other points used to measure the length of the dendritic spine.
- Imaris software gives info like volume, area, length or centre of mass.
- Sometimes incorrectly segmented dendritic spines by the software (24%).

⇒ Correct and complete each spine using a computer vision algorithm.
Clustering of spines

- **D2 shape descriptor**, in which the distance between two random points in the surface is measured and then represented in a histogram [Osada et al., 2001] implying 5000 variables.

- Dissimilarity between objects using the Kullback-Leibler divergence.

- 10 clusters (decided from Silhouette, Davies-Bouldin and Dunn’s measures).

Groups 1 and 3 are medium size and with less defined neck; Groups 4 and 5 are big...

To be improved since having or not neck is not distinguishing, size is more important than shape, and doesn’t describe orientation within the dendrite.
Introduction

Standard machine learning methods
- Clustering for neuron classification
- Clustering of spines
- Knowledge discovery in Alzheimer’s disease
- Supervised classification of neurons
- Supervised classification of dementia development in Parkinson’s
- Supervised classification for outcome prediction of epilepsy surgery

Advanced statistical and machine learning methods
- EM based subspace clustering for discovering new types of neurons
- Bayesian classifiers for probabilistic class neuron
- Multi-dimensional classification for EQ-5D health states in Parkinson
- Computer simulation of dendritic morphology with Bayesian networks
- Spatial point processes for distribution of synapses
- Local regularized regression to decode cognitive states in fMRI experiments

Optimization
- Categorizing Parkinson disease based on non-motor symptoms
- EDAs for parameter optimization in compartmental model
- Multi-objective GAs for brain networks
- Classifying MEG data with lasso logistic regression and EDAs

Conclusions
Alzheimer’s disease

- Neurological disorder primarily affecting the elderly that manifests through memory disorders, cognitive decline and loss of autonomy

Alois Alzheimer (1864-1915)

- In 2006, **26.6 million** sufferers worldwide. Alzheimer’s is predicted to affect **1 in 85** people globally by 2050
- Every **70 seconds**, someone develops Alzheimer’s
- Alzheimer’s is the **seventh**-leading cause of death
- AD is one of the **most costly** diseases to society
Knowledge discovery in Alzheimer’s disease

Alzheimer’s disease and DNA microarrays

- Simultaneously measure thousands of gene expression levels. The most common although recent (2000) technology for AD studies.

Idea [Small et al., 2005]: generate microarray data selectively from the brain site most vulnerable to AD to maximize expression differences between AD and controls.

- Generate data also from a neighboring region relatively resistant to AD.

⇒ Hippocampus is particularly vulnerable to AD (postmortem studies, fMRI data in living subjects).

⇒ Regions: entorhinal cortex (EC) forms the main input to the hippocampus, the most vulnerable; dentate gyrus (DG), the most resistant.
Knowledge discovery in Alzheimer’s disease

Alzheimer’s disease and DNA microarrays

- 6 AD brains + 6 controls, obtained at autopsy ⇒ 24 tissue samples and 7,610 variables (probes)
- [Small et al., 2005]: Several ANOVA using group (AD vs. control), age (33-98 years old) and expression levels (EC/DG)

⇒ Re-analyze the data differently to gain robustness (small sample size!)
⇒ Find out explicit new (or validate old) biological relationships and genes not previously reported
Knowledge discovery in Alzheimer’s disease

Armananzas R., Larrañaga P., Bielza C. (2011)
Ensemble transcript interaction networks: A case study on Alzheimer’s disease
Computer Methods and Programs in Biomedicine
submitted
Knowledge discovery in Alzheimer’s disease

Reliable-$k$DB classifier with robust gene interactions

- Learn a **Bayesian network classifier**. We use **$k$DB** structures with **at most $k$ parents** (excluding the class)
- Search of structure based on the (conditional) **mutual information metric**

![Diagram of Bayesian network](image)

- Induce many **$k$DB** by a resampling method (**bootstrap**) with an inner FSS
- Output a network with those arcs above a reliability threshold $t$: arcs occurring $\geq t$ times are retained [Armañanzas et al., 2008]
- Different confidence level $t$, different model $G_t$ (from simple structures with a few arcs when $t \uparrow$, to dense graphs when $t \downarrow$): $G_1 \supseteq G_2 \supseteq \ldots$
- Approach is a **consensus** feature selection on the final gene interaction network

C. Bielza, P. Larrañaga
Reliable-\(k\)DB classifier – An example

\[ G_t \]

\( X_1 \) connected to \( X_2 \) with a weight of 800.

Graph showing time points: 800, 798, 790, 700, 600, 500, 400, 100, 1.
Reliable-$k$DB classifier – An example

![Diagram showing neurons and spines](image_url)
Reliable-\(k\)DB classifier – An example
Reliable-\(k\)-DB classifier – An example

\[ G_t \]
Reliable-\(k\)DB classifier – An example
Reliable-$k$DB classifier – An example
Reliable-\(k\)DB classifier – An example
Reliable-$k$DB classifier – An example
Reliable-$k$DB classifier – An example
Knowledge discovery in Alzheimer’s disease

**Gene-interaction networks**

- 2 different analyses:
  - EC-AD vs. EC-Control (12 samples, binary supervised classif problem), and
  - EC-AD vs. EC-Control vs. DG-AD vs. DG-Control (24 samples, multiclass)

- $k = 4$, $B = 10,000$ (to avoid local optima), equal width discretization into 3 intervals (up-regulated, down-regulated, baseline or null activity)

- $t = 1,000$ (also could be the 0.999 quantile from the empirical distribution) guaranteeing few false positives

- In the 1st analysis, 2 disconnected graphs. Genes related to neurological diseases: Huntington, Parkinson, bipolar disorder, depression...

- Also, genes not directly related to AD by previous studies → New findings in the pathogenesis and development of AD
Outline

1 Introduction

2 Standard machine learning methods
   - Clustering for neuron classification
   - Clustering of spines
   - Knowledge discovery in Alzheimer’s disease
   - Supervised classification of neurons
     - Supervised classification of dementia development in Parkinson’s
     - Supervised classification for outcome prediction of epilepsy surgery

3 Advanced statistical and machine learning methods
   - EM based subspace clustering for discovering new types of neurons
   - Bayesian classifiers for probabilistic class neuron
   - Multi-dimensional classification for EQ-5D health states in Parkinson
   - Computer simulation of dendritic morphology with Bayesian networks
   - Spatial point processes for distribution of synapses
   - Local regularized regression to decode cognitive states in fMRI experiments

4 Optimization
   - Categorizing Parkinson disease based on non-motor symptoms
   - EDAs for parameter optimization in compartmental model
   - Multi-objective GAs for brain networks
   - Classifying MEG data with lasso logistic regression and EDAs

5 Conclusions
Neurons are more varied than cells in any other part of the body.

Helpful to have a supervised classification model to automatically classify neurons (first step to understand neural circuits).

2 principal neuronal types of the cerebral cortex:

- **Pyramidal cells** \( \approx 80\% \) of the total population
- **Interneurons** \( \approx 20\% \) of the total population: a heterogeneous group

Previous work to classify cortical neurons based on unsupervised clustering (exploratory; w/o using prior info on the potential outcomes). Also, after PCA or with those vars highly correlated with PCs.
Supervised classification of neurons

Comparison Between Supervised and Unsupervised Classifications of Neuronal Cell Types: A Case Study

Luis Guerra, Laura M. McGarry, Víctor Robles, Concha Bielza, Pedro Larrañaga, Rafael Yuste

1 Departamento de Inteligencia Artificial, Facultad de Informática, Universidad Politécnica de Madrid, Spain
2 HHMI, Department of Biological Sciences, Columbia University, New York
3 Departamento de Arquitectura y Tecnología de Sistemas Informáticos, Facultad de Informática, Universidad Politécnica de Madrid, Spain

Received 27 April 2010; accepted 2 June 2010

ABSTRACT: In the study of neural circuits, it becomes essential to discern the different neuronal cell types that build the circuit. Traditionally, neuronal cell types have been classified using qualitative descriptors. More recently, several attempts have been made to classify neurons quantitatively, using unsupervised clustering methods. While useful, these algorithms do not take advantage of previous information known to the investigator, which could improve the classification task. For neocortical GABAergic interneurons, the problem to discern among different cell types is particularly difficult and better methods are needed to perform objective classifications. Here we explore the use of supervised classification algorithms to classify neurons based on their morphological features, using a database of 128 pyramidal cells and 199 interneurons from mouse neocortex. To evaluate the performance of different algorithms we used, as a benchmark, the test to automatically distinguish between pyramidal cells and interneurons, defining “ground truth” by the presence or absence of an apical dendrite. We compared hierarchical clustering with a battery of different supervised classification algorithms, finding that supervised classifications outperformed hierarchical clustering. In addition, the selection of subsets of distinguishing features enhanced the classification accuracy for both sets of algorithms. The analysis of selected variables indicates that dendritic features were most useful to distinguish pyramidal cells from interneurons when compared with somatic and axonal morphological variables. We conclude that supervised classification algorithms are better matched to the general problem of distinguishing neuronal cell types when some information on those cell groups, in our case being pyramidal or interneuron, is known a priori. As a spin-off of this methodological study, we provide several methods to automatically distinguish neocortical pyramidal cells from interneurons, based on their morphologies.

Keywords: supervised; classification; clustering; pyramidal cell; interneuron

C. Bielza, P. Larrañaga

Machine Learning and Optimization in Neuroscience
Supervised classification of neurons

- 327 cells (199 GABAergic interneurons + 128 pyramidal) from mouse neocortex taken from Columbia University
- Presence/absence of an apical dendrite used as “ground truth” for giving labels
- 65 morphological variables (29 dendritic variables, 29 axonal variables, 7 somatic variables) using Neurolucida software (light microscope)

**Algorithms (Weka)**
- Hierarchical clustering (Ward’s method + Euclidean distance; in R)
- Naive Bayes
- Classification tree C4.5
- k-nearest neighbors
- Multi-layer perceptron
- Logistic regression

**Feature subset selection (FSS):**
- Multivariate filtering (correlation-based feature selection (CFS)) vs wrapper
- Forward vs backward vs genetic algorithms
Supervised classification of neurons via supervised and clustering methods

Interneurons (red) and pyramidal (blue). Two major clusters but many misclassifications.
### Accuracy results

<table>
<thead>
<tr>
<th>Method</th>
<th>No FSS</th>
<th>Validation: 10-cv</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hierarchical clustering</td>
<td>59.33</td>
<td>65</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>80.73 ± 10.44</td>
<td>65</td>
</tr>
<tr>
<td>Classification tree</td>
<td>84.40 ± 3.84</td>
<td>65</td>
</tr>
<tr>
<td>5-nearest neighbors</td>
<td>83.18 ± 7.15</td>
<td>65</td>
</tr>
<tr>
<td><strong>Multi-layer perceptron (1)</strong></td>
<td><strong>87.46 ± 9.06</strong></td>
<td><strong>65</strong></td>
</tr>
<tr>
<td>Logistic regression</td>
<td>82.26 ± 7.36</td>
<td>65</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>Forward #</th>
<th>Backward #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hierarchical clustering</td>
<td>77.68</td>
<td>71.25</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>79.82 ± 9.86</td>
<td>79.51 ± 9.74</td>
</tr>
<tr>
<td><strong>Classification tree (2)</strong></td>
<td><strong>82.26 ± 7.17</strong></td>
<td><strong>88.07 ± 6.09</strong></td>
</tr>
<tr>
<td>5-nearest neighbors</td>
<td>83.79 ± 9.55</td>
<td>84.71 ± 6.03</td>
</tr>
<tr>
<td>Multi-layer perceptron</td>
<td>82.57 ± 9.54</td>
<td>87.77 ± 6.36</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>82.26 ± 9.82</td>
<td>85.63 ± 8.56</td>
</tr>
</tbody>
</table>

**MULTIVARIATE FILTERING (CFS)**

<table>
<thead>
<tr>
<th>Method</th>
<th>Forward #</th>
<th>Backward #</th>
<th>Genetic algorithms #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hierarchical clustering</td>
<td>77.68</td>
<td>71.25</td>
<td>79.82</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>79.82 ± 9.86</td>
<td>79.51 ± 9.74</td>
<td>80.43 ± 7.07</td>
</tr>
<tr>
<td>Classification tree (2)</td>
<td><strong>82.26 ± 7.17</strong></td>
<td><strong>88.07 ± 6.09</strong></td>
<td><strong>81.65 ± 7.24</strong></td>
</tr>
<tr>
<td>5-nearest neighbors</td>
<td>83.79 ± 9.55</td>
<td>84.71 ± 6.03</td>
<td>85.01 ± 5.60</td>
</tr>
<tr>
<td>Multi-layer perceptron</td>
<td>82.57 ± 9.54</td>
<td>87.77 ± 6.36</td>
<td>82.26 ± 9.17</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>82.26 ± 9.82</td>
<td>85.63 ± 8.56</td>
<td>83.49 ± 9.45</td>
</tr>
</tbody>
</table>

**WRAPPER**

<table>
<thead>
<tr>
<th>Method</th>
<th>Forward #</th>
<th>Backward #</th>
<th>Genetic algorithms #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive Bayes (3)</td>
<td>87.16 ± 6.34</td>
<td>83.18 ± 9.12</td>
<td>83.49 ± 8.55</td>
</tr>
<tr>
<td>Classification tree (4,5)</td>
<td>86.85 ± 5.29</td>
<td><strong>87.16 ± 5.83</strong></td>
<td><strong>86.85 ± 4.72</strong></td>
</tr>
<tr>
<td>5-nearest neighbors (6,7)</td>
<td>89.30 ± 7.58</td>
<td>86.85 ± 6.26</td>
<td><strong>87.46 ± 5.68</strong></td>
</tr>
<tr>
<td>Multi-layer perceptron (8,9)</td>
<td>88.07 ± 4.99</td>
<td><strong>88.07 ± 8.27</strong></td>
<td><strong>87.46 ± 6.26</strong></td>
</tr>
<tr>
<td>Logistic regression (10,11)</td>
<td>85.63 ± 9.79</td>
<td>84.71 ± 7.54</td>
<td><strong>91.13 ± 5.95</strong></td>
</tr>
</tbody>
</table>

FSS enhanced the accuracy of clustering. Supervised classification outperformed hierarchical clustering. Wrapper generally superior.
Supervised classification of neurons

Comparison of performances: the top models

<table>
<thead>
<tr>
<th></th>
<th>FSS</th>
<th>Search</th>
<th>Algorithm</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No FSS</td>
<td></td>
<td>MLP</td>
<td>0.091</td>
</tr>
<tr>
<td>2</td>
<td>Filter</td>
<td>Backward</td>
<td>C4.5</td>
<td>0.095</td>
</tr>
<tr>
<td>3</td>
<td>Wrapper</td>
<td>Forward</td>
<td>NB</td>
<td>0.095</td>
</tr>
<tr>
<td>4</td>
<td>Wrapper</td>
<td>Forward</td>
<td>5-NN</td>
<td>0.220</td>
</tr>
<tr>
<td>5</td>
<td>Wrapper</td>
<td>Forward</td>
<td>MLP</td>
<td>0.063</td>
</tr>
<tr>
<td>6</td>
<td>Wrapper</td>
<td>Forward</td>
<td>LR</td>
<td>0.053</td>
</tr>
<tr>
<td>7</td>
<td>Wrapper</td>
<td>Backward</td>
<td>C4.5</td>
<td>0.077</td>
</tr>
<tr>
<td>8</td>
<td>Wrapper</td>
<td>Backward</td>
<td>MLP</td>
<td>0.115</td>
</tr>
<tr>
<td>9</td>
<td>Wrapper</td>
<td>Genetic</td>
<td>C4.5</td>
<td>0.052</td>
</tr>
<tr>
<td>10</td>
<td>Wrapper</td>
<td>Genetic</td>
<td>5-NN</td>
<td>0.052</td>
</tr>
<tr>
<td>11</td>
<td>Wrapper</td>
<td>Genetic</td>
<td>LR</td>
<td>—</td>
</tr>
</tbody>
</table>

- Results of **Wilcoxon signed-rank test**. Models not significantly different from the model Wrapper-Genetic-LR.
- An appropriate FSS (**wrapper** in our case) is more important than using a **specific** supervised algorithm.
Supervised classification of neurons

Features chosen (the most commonly selected variables by the 11 winners)

- **Somatic variables**: somatic compactness
- **Axonal variables**: number of axonal Sholl sections, standard deviation of the average axonal segment length, axonal local angle average
- **Dendritic variables**: number of dendritic Sholl sections, ratio of dendritic length to surface area, highest order dendritic segment

Results using separately the somatic, axonal and dendritic subsets of variables

- **Only somatic**: $\sim 60\%$ of accuracy
- **Only axonal**: $\sim 75\%$ of accuracy
- **Only dendritic**: $\sim 85\%$ of accuracy (the most useful to differentiate morphologically pyramidal neurons from interneurons)
Outline

1. Introduction

2. **Standard machine learning methods**
   - Clustering for neuron classification
   - Clustering of spines
   - Knowledge discovery in Alzheimer’s disease
   - Supervised classification of neurons
   - Supervised classification of dementia development in Parkinson’s
   - Supervised classification for outcome prediction of epilepsy surgery

3. **Advanced statistical and machine learning methods**
   - EM based subspace clustering for discovering new types of neurons
   - Bayesian classifiers for probabilistic class neuron
   - Multi-dimensional classification for EQ-5D health states in Parkinson
   - Computer simulation of dendritic morphology with Bayesian networks
   - Spatial point processes for distribution of synapses
   - Local regularized regression to decode cognitive states in fMRI experiments

4. **Optimization**
   - Categorizing Parkinson disease based on non-motor symptoms
   - EDAs for parameter optimization in compartmental model
   - Multi-objective GAs for brain networks
   - Classifying MEG data with lasso logistic regression and EDAs

5. **Conclusions**
Parkinson’s Disease (PD) is one of the most common neurodegenerative diseases; 1% of the population > 60 y.o.

Characterized by a motor disorder but also by cognitive function impairment. Dementia affects ~40% of PD patients.

Distinguishing between dementia and mild cognitive impairment (MCI) is essential on earlier therapeutic intervention to prevent cognitive decline in PD.

Previous MRI studies assumed preselected structure(s) to be investigated, with arbitrary boundaries traced manually or semi-automatically.

- Discern between PD patients cognitively intact, mild cognitive impairment, and dementia.
- Identify the most predictive neuroanatomic biomarkers (to be used for early diagnosis of dementia and MCI).
Predicting dementia development in Parkinson’s disease using Bayesian network classifiers
Psychiatry Research. Neuroimaging
submitted
Supervised classification of dementia development in PD

- **Automatically** identify the most discriminating cerebral regions by combining FSS with Bayesian network classifiers
- **45 PD subjects**: 16 cognitively intact, 15 MCI and 14 dementia
- **Labels** using (slow) neuropsychological tests: *Clinical Dementia Rating Scale* and *Diagnostic and Statistical Manual of Mental Disorders*
- **MRI** from Hospital Santa Creu i Sant Pau (Barcelona)
- Cortical parcellation and subcortical segmentation of images with Freesurfer (uses Surface- and Voxel-Based Morphometry) ⇒ 112 variables on cortical thickness and volume of subcortical structures
- **Algorithms** (Elvira and Weka):
  - Naive Bayes (NB)
  - Selective naive Bayes (FSNB), univariate filter with $I(X_i, C)$
  - CFS-naive Bayes
  - SVM, often used in neuroimaging
- Discretization for Bayesian classifiers by using Fayyad and Irani’s method
Sup. class. of dementia development in PD

Accuracy results and selected features

- **Binary classifiers**: intact vs MCI, MCI vs dementia, intact vs dementia. Also, the 3 classes together
- Validation: 5-cv
- Kruskal-Wallis non-parametric test with $\alpha = 0.05$

<table>
<thead>
<tr>
<th>Classifier</th>
<th>1.Dementia/intact</th>
<th>2.MCI/intact</th>
<th>3.Dementia/MCI</th>
<th>4.Dementia/MCI/intact</th>
</tr>
</thead>
<tbody>
<tr>
<td>NB</td>
<td>93.33± 9.12</td>
<td>86.66± 13.40</td>
<td>96.55± 7.85</td>
<td>64.44±14.48</td>
</tr>
<tr>
<td>FSNB</td>
<td>93.33± 10.66</td>
<td>89.00± 14.48</td>
<td>96.66±10.33</td>
<td>70.00±26.66</td>
</tr>
<tr>
<td>CFS-NB</td>
<td>97.00± 6.74</td>
<td>90.09± 8.40</td>
<td>96.55± 7.85</td>
<td>68.88±16.48</td>
</tr>
<tr>
<td>SVM</td>
<td>96.67±10.82</td>
<td>84.10±15.94</td>
<td>79.31±13.84</td>
<td>62.22±18.59</td>
</tr>
</tbody>
</table>

- FSS improved performance in general. Different variables in each study, automatically identified
  - 1. white matter, hippocampus and entorhinal cortex (reduced in dementia), inferior lateral ventricles (enlargement)
  - 2. brain stem and hippocampus (reduced in MCI)
  - 3. caudate and entorhinal cortex (reduced in dementia), inferior lateral ventricle (enlargement)
  - 4. thalamus, entorhinal cortex, caudal anterior cingulate and fusiform (reduction), inferior lateral ventricle (enlargement)
Selected features

- **Relevant volumes**: enlargement of the lateral inferior ventricles and reduction of hippocampus and white matter.
- Ventricular enlargement already reported before but our variables localize more specifically in the inferior horn (in the temporal lobe).
- Relevant cortical **thickness**: reduction in the entorhinal thickness in the right and left hemispheres.
- Identifying regions of the PD brain using neuroimaging methods improve the understanding of the natural course of dementia and enable early diagnosis of dementia and MCI.
Outline

1. Introduction

2. Standard machine learning methods
   - Clustering for neuron classification
   - Clustering of spines
   - Knowledge discovery in Alzheimer’s disease
   - Supervised classification of neurons
   - Supervised classification of dementia development in Parkinson’s
   - Supervised classification for outcome prediction of epilepsy surgery

3. Advanced statistical and machine learning methods
   - EM based subspace clustering for discovering new types of neurons
   - Bayesian classifiers for probabilistic class neuron
   - Multi-dimensional classification for EQ-5D health states in Parkinson
   - Computer simulation of dendritic morphology with Bayesian networks
   - Spatial point processes for distribution of synapses
   - Local regularized regression to decode cognitive states in fMRI experiments

4. Optimization
   - Categorizing Parkinson disease based on non-motor symptoms
   - EDAs for parameter optimization in compartmental model
   - Multi-objective GAs for brain networks
   - Classifying MEG data with lasso logistic regression and EDAs

5. Conclusions
About 50 million people worldwide have epilepsy.

Chronic and characterized by recurrent seizures: blackouts, involuntary movement, distorted perception.

Such seizures are the external symptom of abnormal neuronal activity in the brain.

30% of epilepsy patients have no control over seizure events even with the best anticonvulsant medications.

TLE, the most common epilepsy of adults.

Brain surgery is considered. However, after surgery, a minor percentage of patients still present occasional seizures and other collateral side-effects.

Machine learning tools revealed personality style as one of the outstanding features to predict outcome of temporal lobe epilepsy surgery

Epilepsia

submitted
### Outcome prediction of epilepsy surgery

Predict whether a patient with TLE will fully recover from epilepsy or not after surgery

- **23 patients** suffering pharmacoresistant TLE (Hospital de la Princesa at Madrid)

<table>
<thead>
<tr>
<th>Clinical features</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male, Female</td>
</tr>
<tr>
<td>Side</td>
<td>Left, Right</td>
</tr>
<tr>
<td>Surgery age</td>
<td>17-32, 33-54</td>
</tr>
<tr>
<td>Onset age</td>
<td>0-1, 2-10, 11-20</td>
</tr>
<tr>
<td>Elapsed time</td>
<td>7-13, 14-19, 20-39</td>
</tr>
<tr>
<td>Type of seizure</td>
<td>Generalized, Partial Complex, Both</td>
</tr>
<tr>
<td>Frequency of seizures</td>
<td>Daily, Weekly, 2-Weekly, 3-Weekly, 4-Weekly, Other</td>
</tr>
<tr>
<td>Febrile episodes</td>
<td>True, False</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Psychological features</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>before and after surgery</td>
<td></td>
</tr>
<tr>
<td>Intelligence quotients</td>
<td>VerbalIQ, PerformanceIQ, TotalIQ</td>
</tr>
<tr>
<td></td>
<td>7 categories: Very Low (0-69)... Very High (≥130)</td>
</tr>
<tr>
<td>Logical/visual memory on:</td>
<td></td>
</tr>
<tr>
<td>immediate recalls</td>
<td>MlogI, MvisI</td>
</tr>
<tr>
<td>delayed recalls</td>
<td>MlogII, MvisII</td>
</tr>
<tr>
<td></td>
<td>5 groups: Low (50-65)... High (≥131)</td>
</tr>
<tr>
<td>From the Rorschach test:</td>
<td></td>
</tr>
<tr>
<td>indices of schizophrenia</td>
<td>sczi, +, —</td>
</tr>
<tr>
<td>social ability</td>
<td>cdi, +, —</td>
</tr>
<tr>
<td>depression</td>
<td>depi</td>
</tr>
<tr>
<td>personality style</td>
<td>Introversive, Extroversive, Ambitent</td>
</tr>
<tr>
<td>Class=Engel scale</td>
<td>Seizure free, Only improvement (after surgery)</td>
</tr>
</tbody>
</table>

- Many **neuropsychological** profiles, that may change after surgery, but **not** previously evaluated in detail
- Missing values imputed using the mode conditioned to the class variable
Recovery from epilepsy or not after surgery

- **Algorithms**: Naive Bayes, Logistic regression and $k$-nearest neighbors
- 19 pre-surgical variables
- Validation: leave-one-out cross-validation
- Small size ⇒ Robust and reliable FSS is important to avoid overfitting

- Produce 1000 intermediate datasets $D_i$ by random stratified resampling with replacement
- In each $D_i$, select features (backwards wrapper FSS (LOOCV))
Outcome prediction of epilepsy surgery

Recovery from epilepsy or not after surgery

- **Algorithms**: Naive Bayes, Logistic regression and $k$-nearest neighbors
- 19 pre-surgical variables
- Validation: leave-one-out cross-validation
- Small size $\Rightarrow$ Robust and reliable FSS is important to avoid overfitting
- Produce 1000 intermediate datasets $D_i$ by random stratified resampling with replacement
- In each $D_i$, select features (backwards wrapper FSS (LOOCV))
Outcome prediction of epilepsy surgery

Relevance analysis. # times each feature has been included in the model

<table>
<thead>
<tr>
<th>Ranking of features</th>
<th>Naïve Bayes freq.</th>
<th>Logistic regression freq.</th>
<th>k-NN freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>P.Style 958</td>
<td>Side 896</td>
<td>Side 891</td>
</tr>
<tr>
<td>2</td>
<td>Side 899</td>
<td>PIQ 885</td>
<td>SeizureFreq 870</td>
</tr>
<tr>
<td>3</td>
<td>PIQ 867</td>
<td>P.Style 875</td>
<td>P.Style 857</td>
</tr>
<tr>
<td>4</td>
<td>SeizureFreq 678</td>
<td>VIQ 862</td>
<td>VIQ 849</td>
</tr>
<tr>
<td>5</td>
<td>SurgeryAge 603</td>
<td>Depi 804</td>
<td>IQ 829</td>
</tr>
<tr>
<td>6</td>
<td>Sczi 596</td>
<td>IQ 772</td>
<td>OnsetAge 800</td>
</tr>
<tr>
<td>7</td>
<td>Gender 477</td>
<td>SurgeryAge 769</td>
<td>Sczi 773</td>
</tr>
<tr>
<td>8</td>
<td>Depi 446</td>
<td>Sczi 747</td>
<td>PIQ 765</td>
</tr>
<tr>
<td>... 19</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C. Bielza, P. Larrañaga
Machine Learning and Optimization in Neuroscience
Final model: iteratively add features one at a time following the previous ranking... until all variables (19) are included.

Best performance:

<table>
<thead>
<tr>
<th></th>
<th>Acc.</th>
<th>AUC</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>NB</td>
<td>89.47%</td>
<td>0.9285</td>
<td>3</td>
</tr>
<tr>
<td>LReg</td>
<td>89.47%</td>
<td>0.9428</td>
<td>3</td>
</tr>
<tr>
<td>kNN</td>
<td>89.47%</td>
<td>0.8071</td>
<td>4</td>
</tr>
</tbody>
</table>

Much worse with *all* vars.
Selected features

- Very relevant features: P.Style, PIQ and Side. Also significant with Wilcoxon signed-rank test of each feature given $C = c$:
  - P. Style ($p = .0134$), Side ($p = .0433$), PIQ ($p = .0492$)

- E.g. PIQ correlated with cortical grey matter thickness in some regions of the temporal cortex (and TLE patients have reduced cortical thickness of all lobes)

⇒ Include these as standard tests in the clinical practice to select an epileptic candidate to surgery: non expensive, relatively easy to apply and non invasive

- No differences between psychological features before and after surgery in both class groups (i.e. they are unchanged no matter of the surgical results)

- Clustering (EM) with post-psychological features in 2 groups (good): post-psychological tests linked to output of the surgery
Outline

1. Introduction
2. Standard machine learning methods
   - Clustering for neuron classification
   - Clustering of spines
   - Knowledge discovery in Alzheimer’s disease
   - Supervised classification of neurons
   - Supervised classification of dementia development in Parkinson’s
   - Supervised classification for outcome prediction of epilepsy surgery
3. Advanced statistical and machine learning methods
   - EM based subspace clustering for discovering new types of neurons
   - Bayesian classifiers for probabilistic class neuron
   - Multi-dimensional classification for EQ-5D health states in Parkinson
   - Computer simulation of dendritic morphology with Bayesian networks
   - Spatial point processes for distribution of synapses
   - Local regularized regression to decode cognitive states in fMRI experiments
4. Optimization
   - Categorizing Parkinson disease based on non-motor symptoms
   - EDAs for parameter optimization in compartmental model
   - Multi-objective GAs for brain networks
   - Classifying MEG data with lasso logistic regression and EDAs
5. Conclusions
Outline

1. Introduction

2. Standard machine learning methods
   - Clustering for neuron classification
   - Clustering of spines
   - Knowledge discovery in Alzheimer's disease
   - Supervised classification of neurons
   - Supervised classification of dementia development in Parkinson's
   - Supervised classification for outcome prediction of epilepsy surgery

3. Advanced statistical and machine learning methods
   - EM based subspace clustering for discovering new types of neurons
   - Bayesian classifiers for probabilistic class neuron
   - Multi-dimensional classification for EQ-5D health states in Parkinson
   - Computer simulation of dendritic morphology with Bayesian networks
   - Spatial point processes for distribution of synapses
   - Local regularized regression to decode cognitive states in fMRI experiments

4. Optimization
   - Categorizing Parkinson disease based on non-motor symptoms
   - EDAs for parameter optimization in compartmental model
   - Multi-objective GAs for brain networks
   - Classifying MEG data with lasso logistic regression and EDAs

5. Conclusions
Since original findings of Santiago Ramón y Cajal, the scientific community is still missing a general catalog of accepted neuron types and names.

The Petilla Terminology by Ascoli et al. (2008)
J.H. Jackson (1874) in *On Classification and Methods of Investigation* defines two types of classification:

- **empirical**: practical purpose (gardener)
- **scientific**: better organization of existing knowledge (botanist)

In neurology, the scientific classification takes into account: morphological, physiological and molecular features.
A gardener classification (not a botanist), coarser and practical

Towards a consensus in naming neocortical neurons
- Most neuroscientists agree with terms like pyramidal, nonpyramidal, interneuron, chandelier (clear morphological attributes)
- Other common names like as double bouquet, basket, Martinotti lack a definition and are used inconsistently

The goal is to establish a list of terms used by all researchers to distinguish neurons with a utilitarian purpose
A gardener neuroclassification

- Classifications of **320 interneurons** given by **42 relevant experts**
- [http://cajalbbp.cesvima.upm.es/gardenerclassification/](http://cajalbbp.cesvima.upm.es/gardenerclassification/)
Discovering new classes

10 types of neurons

<table>
<thead>
<tr>
<th>cell</th>
<th>( X_1 )</th>
<th>( X_{2885} )</th>
<th>( E_1 )</th>
<th>( E_{42} )</th>
<th>( C )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.1</td>
<td>6.6</td>
<td>1</td>
<td>6</td>
<td>4-7-0-10-2-4-3-4-4-4</td>
</tr>
<tr>
<td>2</td>
<td>3.7</td>
<td>7.7</td>
<td>2</td>
<td>2</td>
<td>6-2-1-1-1-4-8-10-9-0</td>
</tr>
<tr>
<td>3</td>
<td>5.9</td>
<td>9.2</td>
<td>1</td>
<td>7</td>
<td>10-10-1-1-0-0-0-2-8-10</td>
</tr>
<tr>
<td>4</td>
<td>11.2</td>
<td>10.1</td>
<td>4</td>
<td>7</td>
<td>6-4-4-2-10-1-1-5-5</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>240</td>
<td>13.6</td>
<td>5.7</td>
<td>7</td>
<td>2</td>
<td>3-2-5-6-4-2-2-2-4-12</td>
</tr>
<tr>
<td>241</td>
<td>8.9</td>
<td>8.8</td>
<td>5</td>
<td>8</td>
<td>7-3-2-2-4-2-10-10-1-1</td>
</tr>
</tbody>
</table>

- Some of the 10 types of neurons (arcade (1), Cajal-Retzius (2), Chandelier (3), common basket (4), horse-tail (5), large basket (6), Martinotti (7), neurogliaform (8), common type (9), other (10)) are selected by a small number of experts.
- Establish a threshold in frequency to select the types to be considered.
- Only some of the 10 types will be selected.
- Goal: develop a semi-supervised algorithm (subspace + EM) for discover new types of neurons (check if the new types agree with some of the original).
GABAergic interneurons of the cerebral cortex: A gardener's classification and nomenclature approach
*Nature Neuroscience*
submitted

Guerra L., Robles V., Bielza C., Larrañaga P. (2011)
A subspaces model-based clustering algorithm for semi-supervised learning
*IEEE Transactions on Pattern Analysis and Machine Intelligence*
submitted
A subspaces EM algorithm for class discovery in semi-supervised learning

Notation

- A set of instances \( \mathcal{X} = \{ \mathbf{x}_1, \ldots, \mathbf{x}_N \} \) with \( \mathbf{x}_i \in \mathbb{R}^F \)
- \( \mathcal{X} = \mathcal{X}^L \cup \mathcal{X}^U \), with \( \mathcal{X}^L = \{ \mathbf{x}_1, \ldots, \mathbf{x}_L \} \) the subset of instances with associated known class label and \( \mathcal{X}^U = \{ \mathbf{x}_{L+1}, \ldots, \mathbf{x}_N \} \) those with unknown labels
- Latent variables set \( \mathcal{Z} = \mathcal{Z}^L \cup \mathcal{Z}^U = \{ \mathbf{z}_1, \ldots, \mathbf{z}_L \} \cup \{ \mathbf{z}_{L+1}, \ldots, \mathbf{z}_N \} \)
  - \( C \) denotes the number of known classes, \( K \) the number of total final classes
  - \( \mathbf{z}_i = (z_{i1}, \ldots, z_{iC}, z_{iC+1}, \ldots, z_{iK}) \)
  - For \( \mathbf{z}_i \in \mathcal{Z}^L \) we have \( z_{im} = 1 \) if instance \( i \) belongs to component \( m \) and with all other elements \( z_{im'} = 0, \forall \ m' \neq m \)
  - This restriction does not apply to \( \mathbf{z}_i \in \mathcal{Z}^U \)
- The feature relevance for each component of the mixture is indicated in the set \( \mathcal{V} = \{ \mathbf{v}_1, \ldots, \mathbf{v}_K \} \), being \( \mathbf{v}_m = (v_{m1}, \ldots, v_{mF}) \) with \( v_{mj} = 1 \) if feature \( j \) is relevant to component \( m \) and \( v_{mj} = 0 \) otherwise
  - For each component and feature, \( \rho_{mj} = p(v_{mj} = 1) \), the probability that feature \( j \) is relevant to component \( m \)
A subspaces EM algorithm for class discovery in semi-supervised learning

**Basic theory of finite mixture models**

- **Mixture model with** $K$ **components**: $p(x_i | \Theta) = \sum_{m=1}^{K} \pi_m p(x_i | \theta_m)$
  
  with $\pi_m \geq 0$, $\sum_{m=1}^{K} \pi_m = 1$, and $\Theta = \{\theta_1, \ldots, \theta_K, \pi_1, \ldots, \pi_K\}$

- **Log-likelihood function**: $\log L(\Theta | \mathcal{X}) = \sum_{i=1}^{N} \log(\sum_{m=1}^{K} \pi_m p(x_i | \theta_m))$
  
  The maximization of this log-likelihood function is difficult due to the summation over the components is inside the logarithm function

- **Complete-data log-likelihood function**:
  
  $\log L(\Theta | \mathcal{X}, \mathcal{Z}) = \sum_{i=1}^{N} \sum_{m=1}^{K} z_{im}(\log \pi_m + \log p(x_i | \theta_m))$

  Its maximization is straightforward because the summation is out of the logarithm. But, the latent variables are unknown

- **EM-algorithm** is used as an iterative algorithm to estimate the parameters that maximize the expectation of the complete-data log-likelihood:
  
  - **E-step**: $Q(\Theta, \Theta^{t-1}) = \sum_{\mathcal{Z}} p(\mathcal{Z} | \mathcal{X}, \Theta^{t-1}) \log p(\mathcal{X}, \mathcal{Z} | \Theta)$
  
  - **M-step**: $\Theta^t = \arg \max_{\Theta} Q(\Theta, \Theta^{t-1})$
A subspaces EM algorithm for class discovery in semi-supervised learning

**Our approach**

- **Log-likelihood with the relevance of each variable to each component:**
  \[ \log L(\Theta | \mathcal{X}) = \sum_{i=1}^{N} \log \sum_{m=1}^{K} \pi_m \prod_{j=1}^{F} (\rho_{mj} p(x_{ij} | \theta_{mj}) + (1 - \rho_{mj}) p(x_{ij} | \lambda_{mj})) \]
  where \( \theta_{mj} \) (\( \lambda_{mj} \)) parameters for the density function if attribute \( j \) is (not) relevant to component \( m \)

- **Complete-data log-likelihood function with two latent variables:**
  \[ \log L(\Theta | \mathcal{X}, \mathcal{Z}, \mathcal{V}) = \log \prod_{i=1}^{N} \prod_{m=1}^{K} (\pi_m)^{z_{im}} \prod_{j=1}^{F} (\rho_{mj} p(x_{ij} | \theta_{mj}) + (1 - \rho_{mj}) p(x_{ij} | \lambda_{mj}))^{v_{mj}} \]
  where \( z_{im} \) and \( v_{mj} \) must be estimated

- **Expectation of the complete-data log-likelihood function with subspaces**
  \[ Q(\Theta, \Theta^{t-1}) = \sum_{\mathcal{Z}} \sum_{\mathcal{V}} p(\mathcal{Z}, \mathcal{V} | \mathcal{X}, \Theta^{t-1}) \log p(\mathcal{X}, \mathcal{Z}, \mathcal{V} | \Theta) \]

- **Expectation of the complete-data log-likelihood function with subspaces and a semi-supervised data set:**
  \[ Q(\Theta, \Theta^{t-1}) = E_{\mathcal{Z}, \mathcal{V} | \mathcal{X}^{\mathcal{L}}, \Theta^{t-1}} [\log L_1(\Theta | \mathcal{X}^{\mathcal{L}}, \mathcal{Z}^{\mathcal{L}}, \mathcal{V})] + E_{\mathcal{Z}, \mathcal{V} | \mathcal{X}^{\mathcal{U}}, \Theta^{t-1}} [\log L_2(\Theta | \mathcal{X}^{\mathcal{U}}, \mathcal{Z}^{\mathcal{U}}, \mathcal{V})] \]

- **Selection of the final number of clusters:**
  - Bottom-up approach starting at \( C \) and using a greedy forward each that increases one cluster at a time
  - Comparison between two models with different number of components via the Bayesian information criterion (BIC)

- **Developed for univariate Gaussian:**
  \( \theta_{mj} = (\mu_{\theta_{mj}}, \sigma^2_{\theta_{mj}}) \) and \( \lambda_{mj} = (\mu_{\lambda_{mj}}, \sigma^2_{\lambda_{mj}}) \)
117 neurons, 57 features, 4 classes of neurons

### one class of neurons hidden

<table>
<thead>
<tr>
<th></th>
<th>arcade</th>
<th>common basket</th>
<th>horse-tail</th>
<th>Martinotti</th>
</tr>
</thead>
<tbody>
<tr>
<td>arcade</td>
<td>16</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>common basket</td>
<td>0</td>
<td>30</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>horse-tail</td>
<td>0</td>
<td>0</td>
<td>44</td>
<td>0</td>
</tr>
<tr>
<td>Martinotti</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>arcade</th>
<th>common basket</th>
<th>horse-tail</th>
<th>Martinotti</th>
</tr>
</thead>
<tbody>
<tr>
<td>arcade</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>common basket</td>
<td>14</td>
<td>3</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>horse-tail</td>
<td>0</td>
<td>0</td>
<td>44</td>
<td>0</td>
</tr>
<tr>
<td>Martinotti</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>arcade</th>
<th>common basket</th>
<th>horse-tail</th>
<th>Martinotti</th>
</tr>
</thead>
<tbody>
<tr>
<td>arcade</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>common basket</td>
<td>0</td>
<td>30</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>horse-tail</td>
<td>1</td>
<td>0</td>
<td>43</td>
<td>0</td>
</tr>
<tr>
<td>Martinotti</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>arcade</th>
<th>common basket</th>
<th>horse-tail</th>
<th>Martinotti</th>
</tr>
</thead>
<tbody>
<tr>
<td>arcade</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>common basket</td>
<td>0</td>
<td>30</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>horse-tail</td>
<td>0</td>
<td>0</td>
<td>44</td>
<td>0</td>
</tr>
<tr>
<td>Martinotti</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>18</td>
</tr>
</tbody>
</table>
Outline

1. Introduction

2. Standard machine learning methods
   - Clustering for neuron classification
   - Clustering of spines
   - Knowledge discovery in Alzheimer’s disease
   - Supervised classification of neurons
   - Supervised classification of dementia development in Parkinson’s
   - Supervised classification for outcome prediction of epilepsy surgery

3. Advanced statistical and machine learning methods
   - EM based subspace clustering for discovering new types of neurons
   - Bayesian classifiers for probabilistic class neuron
   - Multi-dimensional classification for EQ-5D health states in Parkinson
   - Computer simulation of dendritic morphology with Bayesian networks
   - Spatial point processes for distribution of synapses
   - Local regularized regression to decode cognitive states in fMRI experiments

4. Optimization
   - Categorizing Parkinson disease based on non-motor symptoms
   - EDAs for parameter optimization in compartmental model
   - Multi-objective GAs for brain networks
   - Classifying MEG data with lasso logistic regression and EDAs

5. Conclusions
<table>
<thead>
<tr>
<th>cell</th>
<th>( X_1 )</th>
<th>...</th>
<th>...</th>
<th>...</th>
<th>( X_{2885} )</th>
<th>( E_1 )</th>
<th>...</th>
<th>...</th>
<th>...</th>
<th>( E_{42} )</th>
<th>( C )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.1</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>6.6</td>
<td>trans</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>intra</td>
<td>38-4</td>
</tr>
<tr>
<td>2</td>
<td>3.7</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>7.7</td>
<td>intra</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>trans</td>
<td>24-18</td>
</tr>
<tr>
<td>3</td>
<td>5.9</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>9.2</td>
<td>intra</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>intra</td>
<td>35-7</td>
</tr>
<tr>
<td>4</td>
<td>11.2</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>10.1</td>
<td>intra</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>intra</td>
<td>10-32</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>240</td>
<td>13.6</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>5.7</td>
<td>intra</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>intra</td>
<td>3-39</td>
</tr>
<tr>
<td>241</td>
<td>8.9</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>8.8</td>
<td>trans</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>intra</td>
<td>28-14</td>
</tr>
</tbody>
</table>

C. Bielza, P. Larrañaga

Machine Learning and Optimization in Neuroscience
López-Cruz P.L., Bielza C., Larrañaga P. (2011)
Bayesian classifiers for probabilistic labels. A case study in neuron classification
in preparation
Imprecise and/or uncertain labels

Supervised classification with probabilistic labels

- \( n \) predictive variables \( \mathbf{X} = (X_1, \ldots, X_n) \) and a class variable \( C \)

- In a regular supervised classification setting: learning set \( \mathcal{D} = \{(x_1, y_1), \ldots, (x_N, y_N)\} \), where \( x_i \) are the values of the predictive variables for instance \( i \), and \( y_i \in \text{val}(C) \) is the class label for instance \( i \)

- However in some real world situations the information about the class labels is imprecise and/or uncertain, and modeled as a basic belief assignment (bba)

- In the gardener neuroclassificator approach we have a probability distribution over the class domain for each instance \( i \): \( y_i = p(Y_i = c_k), k = 1, \ldots, K \), where \( K \) is the cardinality of \( \text{val}(C) \), i.e., the number of class labels

- It is a special case of the more general problem solved by Denoeux (2011) providing a solution that is generalization of supervised classification, unsupervised classification (clustering), semi-supervised clustering (labeled and unlabeled instances) and partially supervised classification
Generalized likelihood criterion

Côme et al. (2009)

- Fit the mixture model \( f(x) = \sum_{k=1}^{K} \pi_k f_{X|k}(x; \theta_k) \) where \( \pi_k \) is the probability of each class: \( \pi_k = p(Y = c_k) \). We need to estimate the parameter vector \( \Psi = (\pi_1, \ldots, \pi_K, \theta_1, \ldots, \theta_K) \) given a dataset \( D = \{(x_1, m_1), \ldots, (x_N, m_N)\} \), where \( m_i \) is the bba encoding the class information about instance \( i \).

- The generalized log-likelihood criterion computes the logarithm of the conditional plausability of \( \Psi \) given \( D \):

\[
\mathcal{L}(\Psi|D) = \ln(p|^{\Psi}(\Psi|D)) = \sum_{i=1}^{N} \ln \left( \sum_{k=1}^{K} pl_{ik}\pi_k f_{X|k}(x; \theta_k) \right) + \nu,
\]

where \( pl_{ik} \) are the plausabilities of each class \( c_k \) for each sample \( i \) according to soft labels \( m_i \) (\( pl_{ik} = pl_i^C(\{c_k\}) \)), and \( \nu \) is a constant independent of \( \Psi \).

- The generalized log-likelihood can be decomposed in

\[
\mathcal{L}(\Psi|D) = Q(\Psi, \Psi^{(q)}) - H(\Psi, \Psi^{(q)}),
\]

\[
Q(\Psi, \Psi^{(q)}) = \sum_{i=1}^{N} \sum_{k=1}^{K} t_{ik}^{(q)} \ln(pl_{ik}\pi_k f_{X|k}(x; \theta_k))
\]

\[
H(\Psi, \Psi^{(q)}) = \sum_{i=1}^{N} \sum_{k=1}^{K} t_{ik}^{(q)} \ln(t_{ik}).
\]
The **EM algorithm** can be used to find a (local) optimum of $\mathcal{L}(\Psi | \mathcal{D})$ for Gaussian mixture models:

- **Expectation:**
  \[
  t_{ik}^{(q)} = \frac{p_l i_k \pi_k f_{x|k}(x_i; \mu_k, \Sigma_k)}{\sum_{k'=1}^{K} p_l i_{k'} \pi_{k'} f_{x|k'}(x_i; \mu_{k'}, \Sigma_{k'})}
  \]

- **Maximization:**
  \[
  \pi_k^{(q+1)} = \frac{1}{N} \sum_{i=1}^{N} t_{ik}^{(q)}
  \]
  \[
  \mu_k^{(q+1)} = \frac{1}{\sum_{i=1}^{N} t_{ik}^{(q)}} \sum_{i=1}^{N} t_{ik}^{(q)} x_i
  \]
  \[
  \Sigma_k^{(q+1)} = \frac{1}{\sum_{i=1}^{N} t_{ik}^{(q)}} \sum_{i=1}^{N} t_{ik}^{(q)} (x_i - \mu_k^{(q+1)}) (x_i - \mu_k^{(q+1)})'
  \]
Results on classification error

- MM PLEM
- NB PLEM
- AODE PLEM
- MM EM
- NB EM
- AODE EM

Results

C. Bielza, P. Larrañaga

Machine Learning and Optimization in Neuroscience
Results on mean squared error

- MM PLEM
- NB PLEM
- AODE PLEM
- MM EM
- NB EM
- AODE EM

Mean squared error vs FSS Method

C. Bielza, P. Larrañaga
Machine Learning and Optimization in Neuroscience
Outline

1. **Introduction**

2. **Standard machine learning methods**
   - Clustering for neuron classification
   - Clustering of spines
   - Knowledge discovery in Alzheimer’s disease
   - Supervised classification of neurons
   - Supervised classification of dementia development in Parkinson’s
   - Supervised classification for outcome prediction of epilepsy surgery

3. **Advanced statistical and machine learning methods**
   - EM based subspace clustering for discovering new types of neurons
   - Bayesian classifiers for probabilistic class neuron
   - **Multi-dimensional classification for EQ-5D health states in Parkinson**
   - Computer simulation of dendritic morphology with Bayesian networks
   - Spatial point processes for distribution of synapses
   - Local regularized regression to decode cognitive states in fMRI experiments

4. **Optimization**
   - Categorizing Parkinson disease based on non-motor symptoms
   - EDAs for parameter optimization in compartmental model
   - Multi-objective GAs for brain networks
   - Classifying MEG data with lasso logistic regression and EDAs

5. **Conclusions**
Several clinical instrument for measure the severity and degree of disability in Parkinson patients

PDQ-39 captures patient’s perception of their illness

PDQ-39 as the most appropriate health related quality of life (QoL) instrument

39 questions, each of them with 5 possible values

PDQ-39 covers 8 dimensions:

1. Mobility
2. Activities of daily living
3. Emotional well-being
4. Stigma
5. Social support
6. Cognitions
7. Communication
8. Bodily discomfort
# Parkinson’s Disease Questionnaire (PDQ-39)

Please complete the following

Due to having Parkinson’s disease, how often during the last month have you....

<table>
<thead>
<tr>
<th></th>
<th>Never</th>
<th>Occasionally</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always or cannot do at all</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Please tick one box for each question*
EuroQol-5 Dimensions (EQ-5D)

- EQ-5D is a standardized measure of health status developed by the EuroQol Group (www.euroqol.org) in order to provide a simple, generic measure of health for clinical and economic appraisal.

- Respondents rate their health status.

- EQ considers 5 items (each of them with 3 possible values):
  1. Mobility
  2. Self-care
  3. Usual activities
  4. Pain-discomfort
  5. Anxiety-depression

- The $243(3^5)$ possible combinations (from (11111) to (33333)) are mapped to an utility index score.

- The utility index score moves into a scale from 0 (dead) to 1 (full health) and is the value that an average person would attach to a given health state.
A European **pharmaceutical company** interested on transforming PDQ-39 values into EQ-5D for Parkinson patients from which only PDQ-39 information is available.

The company owns a data file containing **488 Parkinson patients** PQD39 questionnaire and their corresponding EQ-5D values.

To be used on **cost-effectiveness** or cost-utility studies.

Assignment of **health resources**
Borchani H., Larrañaga P., Bielza C., Gradl B., Martínez P. (2011)
Predicting EQ-5D from PDQ-39 using multi-dimensional Bayesian network classifiers
Quality of Life Research
submitted
Multi-dimensional classification with Bayesian networks

C. Bielza\textsuperscript{a}, G. Li\textsuperscript{b}, P. Larrañaga\textsuperscript{a,*}

\textsuperscript{a} Computational Intelligence Group, Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid, Boadilla del Monte, 28660 Madrid, Spain
\textsuperscript{b} Rega Institute and University Hospitals, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium

\section*{ARTICLE INFO}
\textit{Article history:}
Received 30 July 2010
Revised 2 December 2010
Accepted 21 January 2011
Available online 16 February 2011

\section*{ABSTRACT}
Multi-dimensional classification aims at finding a function that assigns a vector of class values to a given vector of features. In this paper, this problem is tackled by a general family of models, called multi-dimensional Bayesian network classifiers (MBCs). This probabilistic graphical model organizes class and feature variables as three different subgraphs: class variables, feature variables and their interactions.
Methodology based on ...

THE FIFTH EUROPEAN WORKSHOP ON PROBABILISTIC GRAPHICAL MODELS (PGM-2010)

Learning CB-decomposable Multi-dimensional Bayesian Network Classifiers

Hanen Borchani, Concha Bielza and Pedro Larrañaga
Departamento de Inteligencia Artificial, Facultad de Informática
Universidad Politécnica de Madrid, Boadilla del Monte, 28660, Madrid, Spain.
hanen.borchani@upm.es, {mcbielza, pedro.larranaga}@fi.upm.es

Abstract

Multi-dimensional Bayesian network classifiers (MBCs) have been recently introduced to deal with multi-dimensional classification problems where instances are assigned to multiple classes. MBCs have a restricted topology partitioning the set of class and feature variables into three different subgraphs: class subgraph, feature subgraph and bridge subgraph. In this paper, we propose a novel learning algorithm for class-bridge (CB) decomposable MBCs into maximal connected components. Basically, based on a wrapper greedy forward selection approach, the algorithm firstly learns the bridge and feature subgraphs. Then, while the number of components is greater than one and there is an accuracy improvement, it iteratively and sequentially merges together the components, and updates the bridge and feature subgraphs. By learning CB-decomposable MBCs, the computations of MPE are alleviated comparing to general MBCs. Experimental comparison with state-of-the-art algorithms are carried out using synthetic and real-world data sets. The obtained results show the merits of our proposed algorithm.
Methodology based on ...

PROBABILISTIC PROBLEM SOLVING IN BIOMEDICINE
WORKSHOP IN THE 13TH CONFERENCE ON ARTIFICIAL
INTELLIGENCE IN MEDICINE (AIME-2011)

Learning Multi-Dimensional Bayesian Network
Classifiers Using Markov Blankets: A Case Study
in the Prediction of HIV Protease Inhibitors

Hanen Borchani, Concha Bielza, and Pedro Larrañaga

Computational Intelligence Group, Departamento de Inteligencia Artificial, Facultad de Informática, Universidad Politécnica de Madrid, Boadilla del Monte, 28660, Spain.
hanen.borchani@upm.es, mcbielza@fi.upm.es, pedro.larranaga@fi.upm.es

Abstract. Multi-dimensional Bayesian network classifiers (MBCs) are Bayesian network classifiers especially designed to solve multi-dimensional classification problems, where each instance in the data set has to be assigned to one or more class variables. In this paper, we introduce a new method for learning MBCs from data basically based on determining the Markov blanket around each class variable using the HITON algorithm. Our method is applied to the human immunodeficiency virus (HIV) protease inhibitor prediction problem. The experimental study showed promising results in terms of classification accuracy, and we gained insight from the learned MBC structure into the different possible interactions among protease inhibitors and resistance mutations.
Multi-dimensional classification: an extension of multi-label classification

<table>
<thead>
<tr>
<th>$PDQ_1$</th>
<th>$PDQ_2$</th>
<th>...</th>
<th>...</th>
<th>$PDQ_{39}$</th>
<th>$EQ_1$</th>
<th>$EQ_2$</th>
<th>$EQ_3$</th>
<th>$EQ_4$</th>
<th>$EQ_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>...</td>
<td>...</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>...</td>
<td>...</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>...</td>
<td>...</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>...</td>
<td>...</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>...</td>
<td>...</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>...</td>
<td>...</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
Multi-dimensional Bayesian network classifiers (MBC)

**Definition of an MBC**

- The set of vertices $\mathcal{V}$ is partitioned into:
  - $\mathcal{V}_C = \{C_1, \ldots, C_d\}$ of class variables and
  - $\mathcal{V}_X = \{X_1, \ldots, X_m\}$ of feature variables, with $(d + m = n)$
- Class, bridge and feature subgraphs

**Most probable explanation (MPE)**

$$(c_1^*, \ldots, c_d^*) = \max_{c_1, \ldots, c_d} p(C_1 = c_1, \ldots, C_d = c_d | X_1 = x_1, \ldots, X_m = x_m)$$
Three MBC learning algorithms

- **Class-Bridge - Multi-dimensional Bayesian classifier (CB-MBC)** [Borchani et al., 2010]
  - Step 1: Learns an initial bridge subgraph with a number of maximal connected components equal to the number of class variables, then it learns an initial feature subgraph.
  - Step 2: As long as the number of maximal connected components is greater than one and there is an accuracy improvement, it iteratively and sequentially merges together the components and updates the bridge and feature subgraphs.

- **Markov blanket - Independent classifiers (MB-IC)** [Aliferis et al., 2010]
  - Learning independently a classifier for each class variable using the HITON algorithm.

- **Markov blanket - Multi-dimensional Bayesian classifier (MB-MBC)** [Borchani et al., 2011]
  - Step 1: Apply the HITON algorithm to each class variable considering the rest of class variables and all the feature variables as potential nodes in its Markov blanket.
  - Step 2: Specify the directionality over the MBC subgraphs.
Results on 488 Parkinson patients

<table>
<thead>
<tr>
<th></th>
<th>MEAN ACCURACY</th>
<th>GLOBAL ACCURACY</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB-MBC</td>
<td>0.7147 ± 0.0456</td>
<td>0.6498 ± 0.0838</td>
</tr>
<tr>
<td>MB-IC</td>
<td>0.8992 ± 0.0016</td>
<td>0.8759 ± 0.0082</td>
</tr>
<tr>
<td>MB-MBC</td>
<td>0.9847 ± 0.0188</td>
<td>0.9567 ± 0.0539</td>
</tr>
</tbody>
</table>
Structure of the MB-MBC
Outline

1. Introduction

2. Standard machine learning methods
   - Clustering for neuron classification
   - Clustering of spines
   - Knowledge discovery in Alzheimer’s disease
   - Supervised classification of neurons
   - Supervised classification of dementia development in Parkinson’s
   - Supervised classification for outcome prediction of epilepsy surgery

3. Advanced statistical and machine learning methods
   - EM based subspace clustering for discovering new types of neurons
   - Bayesian classifiers for probabilistic class neuron
   - Multi-dimensional classification for EQ-5D health states in Parkinson
   - Computer simulation of dendritic morphology with Bayesian networks
   - Spatial point processes for distribution of synapses
   - Local regularized regression to decode cognitive states in fMRI experiments

4. Optimization
   - Categorizing Parkinson disease based on non-motor symptoms
   - EDAs for parameter optimization in compartmental model
   - Multi-objective GAs for brain networks
   - Classifying MEG data with lasso logistic regression and EDAs

5. Conclusions
Dendritic morphology

- Tree shapes → interconnectivity and functional roles of neurons
- Influenced by complex interaction of synapses, intracellular transport, extracellularly initiated signaling cascades, membrane tension and electrical activity
- How and why vastly different shapes arise are still largely unknown
- Their normal function and why they are often malformed in neurological diseases or under the effects of some drugs (cocaine, morphine)

- Rough groups based on prominent geometrical features. No 2 neurons with the same morphology (location, presence...) → but branching patterns

⇒ Anatomical characterization is statistical in nature
Models and Simulation of 3D Neuronal Dendritic Trees Using Bayesian Networks

Pedro L. López-Cruz · Concha Bielza · Pedro Larrañaga · Ruth Benavides-Piccione · Javier DeFelipe

© Springer Science+Business Media, LLC 2011

Abstract Neuron morphology is crucial for neuronal connectivity and brain information processing. Computational models are important tools for studying dendritic morphology and its role in brain function. We applied a class of probabilistic graphical models called Bayesian networks to generate virtual dendrites from layer III pyramidal neurons from three different regions of the neocortex of the mouse. A set of 41 morphological variables were measured from the 3D reconstructions of real dendrites and their probability distributions used in a machine learning algorithm to induce the model from the data. A simulation algorithm is also proposed to obtain new dendrites by sampling values from Bayesian networks. The main advantage of this approach is that it takes into account and automatically locates the relationships between variables in the data instead of using predefined dependencies. Therefore, the methodology can be applied to any neuronal class while at the same time exploiting class-specific properties. Also, a Bayesian network was defined for each part of the dendrite, allowing the relationships to change in the different sections and to model heterogeneous developmental factors or spatial influences. Several univariate statistical tests and a novel multivariate test based on Kullback–Leibler divergence estimation confirmed that virtual dendrites were similar to real ones. The analyses of the models showed relationships that conform to current neuroanatomical knowledge and support model correctness. At the same time, studying the relationships in the models can help to identify new interactions between variables related to dendritic morphology.

Keywords Pyramidal cells · Virtual dendrites · Morphology simulation · Dendritic structure · Bayesian networks
Our proposal: advantages

Steps

1. Measure from real cells different **parameters** controlling branching behaviour
2. Model these measures as **statistical distributions**
3. Design a **simulation model** to account for all the parameters together and run it
4. Compare formed virtual trees to reality

...with Bayesian networks

- (In)dependences between morphological properties **automatically found** from real data *(vs. prior conditional relationships ad hoc)*
- Possibility of **discovering** non-reported relationships *(vs. biased towards expert knowledge)*
- Model the **joint probability distribution** of all variables *(vs. ≤ trivariate and standard distributions)*
- Reliable evaluation: **statistical tests** to compare original vs. simulated distributions, both uni and multivariate *(vs. on new 1D pars and visual inspection)*
Data: pyramidal neurons

- 3D reconstructions of 90 pyramidal neurons from the mouse neocortex, traced with \textit{Neurolucida} package

- Layer III of different \textbf{cortical regions}: M2, S2, V2L/TeA \Rightarrow 3 databases

- Each basal arbor with 6 (average) main trunks –\textit{dendritic trees}–, each made up of several dendrites

\begin{itemize}
  \item \textbf{Cortex region} \hspace{1cm} Database \hspace{1cm} \# dendr. trees
  \item Motor \hspace{1cm} M2 \hspace{1cm} 104
  \item Somatosensory \hspace{1cm} S2 \hspace{1cm} 103
  \item Lateral visual and association temporal \hspace{1cm} V2L/TeA \hspace{1cm} 156
\end{itemize}

Publicly available at \texttt{http://neuromorpho.org} as part of DeFelipe’s archive (same lab)

C. Bielza, P. Larrañaga

\textbf{Machine Learning and Optimization in Neuroscience}
Morphological parameters

- For each pair of sibling segments (line between two branch points), measure 41 variables.
- Widely used and also new, to capture context influence and neuritic competition.
- Construction variables: define the morphology of a segment (segment length, orientation, bifurcation). **Sampled** by the model to incrementally construct trees.
- Evidence variables: measure the part of the tree previous to a pair of sibling segments (subtree and subdendrite involved). **Measured** during the simulation, used as information to sample construction variables.

---

**Diagram:**

- **A**
  - subdendrite
  - subtree
  - sibling segments
  - neighboring segment
  - parent segment
  - root segment

- **B**
  - subtree height
  - subtree max order
  - subtree min length
  - subtree max length
  - subtree width
  - subtree depth
List of variables

<table>
<thead>
<tr>
<th>No.</th>
<th>Type</th>
<th>Variable</th>
<th>No.</th>
<th>Type</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>E</td>
<td>subtree degree (no. endings)</td>
<td>22</td>
<td>E</td>
<td>neighbor distance</td>
</tr>
<tr>
<td>2</td>
<td>E</td>
<td>subtree no. bifurcations (no. nodes)</td>
<td>23</td>
<td>E</td>
<td>neighbor inclination</td>
</tr>
<tr>
<td>3</td>
<td>E</td>
<td>subtree total length</td>
<td>24</td>
<td>E</td>
<td>neighbor azimuth</td>
</tr>
<tr>
<td>4</td>
<td>E</td>
<td>subtree width</td>
<td>25</td>
<td>E</td>
<td>neighbor extension</td>
</tr>
<tr>
<td>5</td>
<td>E</td>
<td>subtree height</td>
<td>26</td>
<td>E</td>
<td>neighbor angle</td>
</tr>
<tr>
<td>6</td>
<td>E</td>
<td>subtree depth</td>
<td>27</td>
<td>E</td>
<td>parent segment length</td>
</tr>
<tr>
<td>7</td>
<td>E</td>
<td>subtree box volume</td>
<td>28</td>
<td>E</td>
<td>parent segment inclination</td>
</tr>
<tr>
<td>8</td>
<td>E</td>
<td>subtree max distance between nodes</td>
<td>29</td>
<td>E</td>
<td>parent segment azimuth</td>
</tr>
<tr>
<td>9</td>
<td>E</td>
<td>subtree max distance to soma</td>
<td>30</td>
<td>E</td>
<td>root segment length</td>
</tr>
<tr>
<td>10</td>
<td>E</td>
<td>subtree max length</td>
<td>31</td>
<td>E</td>
<td>root segment inclination</td>
</tr>
<tr>
<td>11</td>
<td>E</td>
<td>subtree min length</td>
<td>32</td>
<td>E</td>
<td>root segment azimuth</td>
</tr>
<tr>
<td>12</td>
<td>E</td>
<td>subtree max order</td>
<td>33</td>
<td>E</td>
<td>segment centrifugal order</td>
</tr>
<tr>
<td>13</td>
<td>E</td>
<td>subtree min order</td>
<td>34</td>
<td>C</td>
<td>left segment length</td>
</tr>
<tr>
<td>14</td>
<td>E</td>
<td>subdendrite length</td>
<td>35</td>
<td>C</td>
<td>left segment inclination</td>
</tr>
<tr>
<td>15</td>
<td>E</td>
<td>subdendrite width</td>
<td>36</td>
<td>C</td>
<td>left segment azimuth</td>
</tr>
<tr>
<td>16</td>
<td>E</td>
<td>subdendrite height</td>
<td>37</td>
<td>C</td>
<td>left segment bifurcates</td>
</tr>
<tr>
<td>17</td>
<td>E</td>
<td>subdendrite depth</td>
<td>38</td>
<td>C</td>
<td>right segment length</td>
</tr>
<tr>
<td>18</td>
<td>E</td>
<td>subdendrite box volume</td>
<td>39</td>
<td>C</td>
<td>right segment inclination</td>
</tr>
<tr>
<td>19</td>
<td>E</td>
<td>subdendrite distance to soma</td>
<td>40</td>
<td>C</td>
<td>right segment azimuth</td>
</tr>
<tr>
<td>20</td>
<td>E</td>
<td>subdendrite inclination</td>
<td>41</td>
<td>C</td>
<td>right segment bifurcates</td>
</tr>
<tr>
<td>21</td>
<td>E</td>
<td>subdendrite azimuth</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Variables *discretized* (2-3 values) trying to preserve empirical distributions
Bayesian network learning

Overview of the learning

- Learn and use a BN for each part of the dendritic tree, to allow specific relationships at each part

\[ P(X_1, \ldots, X_{41}) = \prod_{i=1}^{n=41} P(X_i|\Pi_i) \]

\[ \Pi_i = \text{parents of } X_i \text{ in the graph} \]

- Learn the structure via K2 algorithm (a score + search approach)
  - Assume an ordering between nodes (evidence vars before construction vars)
  - Fix an upper bound on the max number of parents for any node (=3)
- Learn the parameters (probabilities) via MLE

\[ P(X_i = x_i|\Pi_i = \pi_i) = \frac{\text{freq}(X_i = x_i, \Pi_i = \pi_i)}{\text{freq}(\Pi_i = \pi_i)} \]
Bayesian networks learnt

For M2 database

- A, B, C, D → root segments, order 1, order 2, > 2 order, resp. Shaded = construction variables

- Found relationships conform to biological knowledge, e.g.
  - Segment length (34, 38) and bifurcation (37, 41) occurrence → more bifurcations close to the soma and shorter segments, whereas segments that do not branch spread away from the soma
Bayesian networks learnt

For M2 database

- A, B, C, D → root segments, order 1, order 2, > 2 order, resp. Shaded = construction variables

- Found relationships conform to biological knowledge, e.g.
  - Subdendrite width (15) and segment bifurcation (37) [wider → doesn't bifurcate] → constrain tree size.
  - Smaller inclination angles (35) for taller subdendrites (16), etc
Simulation of virtual dendritic trees

Procedure (breadth-first way)

1. Generate a root segment
2. Measure evidence variables from the dendritic tree built so far
3. Sample construction variable values from the Bayesian network
4. If a segment bifurcates, consider that the dendrite is still incomplete and go to 2. Else, the dendrite has ended
Model validation

Compare real/simulated prob. distributions ...of variables used in model

- Simulate the **same number of dendritic trees** than in the original database
- **Univariate** statistical tests (compare marginal distributions): KS, Wilcoxon rank-sum, KL (with `bioDist` R package to estimate KL for continuous variables, and bootstrap to estimate percentile 95)
- **Repeat 100 times** to consider statistical variability and perform a **sign test** for each test to check if the number of rejections was significant in the 100 repetitions

![Graph showing number of rejections](image)

- Some rejections in evidence variables in higher orders (high variability)
- No rejections for construction vars
Model validation

Compare real/simulated prob. distributions ...of variables used in model

- **Multivariate** statistical test (compare the joint distribution), used for the first time in this context.
- ...Use **multivariate KL estimator** based on k-NN density estimation [Wang et al., 2006]

\[
\hat{KL}(p||q) = \frac{n}{N_p} \sum_{i=1}^{N_p} \log \frac{\nu_{D_q}(i)}{\rho_{D_p}(i)} + \log \frac{N_q}{N_p - 1}
\]

Datasets \(D_p, D_q\) with \(n\)-dim samples of sizes \(N_p, N_q\),

\(\rho_{D_p}(i)\)=distance from \(x_i \in D_p\) to its NN in \(D_p\), \(\nu_{D_q}(i)\)=distance to its NN in \(D_q\)

» Worse for higher orders (high variability) – perhaps due to the previous variables
Model validation

Compare real/simulated prob. distributions ...of variables NOT used in model

- Measure **12 new** global variables —*emergent*— from real and virtual trees

<table>
<thead>
<tr>
<th>No.</th>
<th>Variable</th>
<th>No.</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>degree (no. endings)</td>
<td>7</td>
<td>dendritic tree depth</td>
</tr>
<tr>
<td>2</td>
<td>no. of bifurcations</td>
<td>8</td>
<td>box volume</td>
</tr>
<tr>
<td>3</td>
<td>total length</td>
<td>9</td>
<td>diameter</td>
</tr>
<tr>
<td>4</td>
<td>mean asymmetry index</td>
<td>10</td>
<td>max distance to soma</td>
</tr>
<tr>
<td>5</td>
<td>dendritic tree width</td>
<td>11</td>
<td>max centrifugal order</td>
</tr>
<tr>
<td>6</td>
<td>dendritic tree height</td>
<td>12</td>
<td>min centrifugal order</td>
</tr>
</tbody>
</table>

- Same statistical tests (uni and multivariate) over trees with the same maximum centrifugal order

- Simulated and real dendrites are similar
- Var7 (dendritic tree depth) rejected: dependent on the resolution of the Z-dim when tracing the neurons, lower than for X and Y ⇒ Z-measures have errors and uncertainty [Steuber et al., 2004], limiting model ability to accurately capture depth
- Equal medians, although wider whiskers and more outliers in simulated
Model validation

Visual inspection by an expert in neuroanatomy

real

virtual
Outline

1. Introduction

2. Standard machine learning methods
   - Clustering for neuron classification
   - Clustering of spines
   - Knowledge discovery in Alzheimer’s disease
   - Supervised classification of neurons
   - Supervised classification of dementia development in Parkinson’s
   - Supervised classification for outcome prediction of epilepsy surgery

3. Advanced statistical and machine learning methods
   - EM based subspace clustering for discovering new types of neurons
   - Bayesian classifiers for probabilistic class neuron
   - Multi-dimensional classification for EQ-5D health states in Parkinson
   - Computer simulation of dendritic morphology with Bayesian networks
   - Spatial point processes for distribution of synapses
   - Local regularized regression to decode cognitive states in fMRI experiments

4. Optimization
   - Categorizing Parkinson disease based on non-motor symptoms
   - EDAs for parameter optimization in compartmental model
   - Multi-objective GAs for brain networks
   - Classifying MEG data with lasso logistic regression and EDAs

5. Conclusions
Spatial distribution of synapses

- **Segment** and 3D reconstruction of synapses in the cerebral cortex from DualBeam electron microscope images, both asymmetric (excitatory) and symmetric (inhibitory)
- Look for spatial patterns (random or specific)
- Distribution of the closest synapse to a given one (important for the neurotransmitter diffusion)

Type I (asym); Type II (sym)
Spatial distribution of synapses

Spatial distribution of synapses in layer III somatosensory cortex via random sequential adsorption model

in preparation
Spatial point processes [Baddeley et al., 2003; Diggle 2003; Illian et al., 2008]

Point pattern
- A point pattern dataset gives the locations of objects/events occurring in a study region
- Points could be trees, nests, earthquake epicentres, petty crimes, galaxies,...
- Situated in a region of the 2D plane, or on the Earths surface, or a 3D volume,...

Point process
- A point process \( X \) is a random set of points; with a random number of points and random locations
- The observed point pattern \( x \) (an unordered set of points) is a realisation of a random point process \( X \)
- It extends throughout the space, but is observed only inside a region \( W \), the sampling window, fixed and known and of any shape
  \[ x = \{x_1, \ldots, x_n\} \quad x_i \in W \]
- Goal: estimate parameters of the distribution of \( X \)
Important concepts

Intensity

- **Expected number of points per unit volume** (frequency)
- May be **constant** (‘uniform’ or ’homogeneous’) across the study volume or may **vary** from location to location (‘non-uniform’ or ‘inhomogeneous’)

Interpoint interaction

- **Stochastic dependence** between the points in a point pattern
- Stronger dependence between points that are **close** to one another is expected
- E.g. space between trees greater than expected at random? (reflecting competition for resources)
Complete spatial randomness (CSR)

Homogeneous spatial Poisson point process, with intensity $\lambda > 0$

- The basic reference or benchmark model of a point process, with the properties:
  - **(P1)** The number of points $N(X \cap B)$ falling in any region $B$ has a Poisson distribution.
  - **(P2)** With mean $\lambda \cdot \text{vol}(B)$ points falling in $B$.
  - **(P3)** For any $B_1, B_2$ disjoint sets, then $N(X \cap B_1)$ and $N(X \cap B_2)$ are independent r.v.
  - **(P4)** Given $n$ points inside region $B$, their locations are i.i.d. and uniformly distributed in $B$.

- P4 represents the general concept of CSR, points with the same propensity to be found at any location regardless of those of other points.
- Sometimes simply called CSR.
- Focus on establishing whether data do not conform to CSR $\Rightarrow$ if completely random, then there is nothing “interesting” happening (points unpredictable, without trend or association with anything else).
State of the art

Statistical methods for spatial point patterns

- Highly developed branch of probability theory for point processes, but the corresponding statistical methodology is relatively underdeveloped. Techniques include:
  - **Summary statistics** (e.g. average distance from a point to its nearest neighbour): literature dominated by these *ad hoc* methods with very little statistical theory to support them.
  - **Comparison to Poisson process**: hypothesis tests to decide whether the point pattern is 'completely random' whether or not this is scientifically relevant.
  - **Modelling** (the last decade): formulate and fit realistic models to point pattern data –a lot of work to be done e.g. in algorithms, model choice, goodness-of-fit.

Software: R packages

The best: **spatstat**
Others: splancs, spatial, ptproc, SSLib
Spatial distribution of synapses

Take into account that a synapse occupies a volume and therefore two synapses cannot overlap

⇒ Model corresponds to a homogeneous Poisson process, where synapses with overlapped volumes are not allowed ⇒ “random sequential adsorption” (RSA) or “simple sequential inhibition” model

Steps:

1. Estimate pdf $f$ of volumes
2. Simulate from RSA model:
   a. Place iteratively and randomly spheres with radii $\sim f$
   b. If a new sphere intersects with an existing one, reject it and generate another with a different center and radius
   c. ...until having the required number of spheres
3. Use simulations to estimate the summary statistics of RSA –F, G and K functions–, and compare with observed, via KS goodness-of-fit tests
4. Finally, model symmetric and asymmetric synapses separately
Spatial distribution of synapses

Data and preprocess

- **10 3D-samples** (mean of 180.33 $\mu m^3$) from 3 Wistar rats in layer III somatosensory cortex

- **ESPINA** used to segment, reconstruct synaptic junctions and obtain relevant data: spatial positions of centroids of junctions, estimation of their sizes (Feret’s diameter=diameter of the smallest sphere containing each)...

C. Bielza, P. Larrañaga

Machine Learning and Optimization in Neuroscience
Spatial distribution of synapses

### Intensities and Volumes

<table>
<thead>
<tr>
<th>Sample</th>
<th>Intensity (syn/(\mu m^3))</th>
<th>Mean distance to the NN (nm)</th>
<th>Mean Feret’s diameter of synaptic junctions (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9857</td>
<td>519.55</td>
<td>377.19</td>
</tr>
<tr>
<td>2</td>
<td>0.6936</td>
<td>594.07</td>
<td>462.18</td>
</tr>
<tr>
<td>3</td>
<td>0.9279</td>
<td>537.43</td>
<td>437.62</td>
</tr>
<tr>
<td>4</td>
<td>1.0088</td>
<td>537.39</td>
<td>414.22</td>
</tr>
<tr>
<td>5</td>
<td>0.9474</td>
<td>597.30</td>
<td>466.03</td>
</tr>
<tr>
<td>6</td>
<td>0.9399</td>
<td>533.21</td>
<td>423.38</td>
</tr>
<tr>
<td>7</td>
<td>0.9881</td>
<td>487.17</td>
<td>397.29</td>
</tr>
<tr>
<td>8</td>
<td>0.7997</td>
<td>568.21</td>
<td>427.90</td>
</tr>
<tr>
<td>9</td>
<td>1.1267</td>
<td>501.38</td>
<td>378.35</td>
</tr>
<tr>
<td>10</td>
<td>1.0178</td>
<td>523.74</td>
<td>405.43</td>
</tr>
</tbody>
</table>

\(f \sim \text{LogNormal distribution for Feret’s diameters:}\)

![Probability Density Function](image)
Spatial distribution of synapses

- **G** CDF of the NN distance for a typical point: $G(r) = P\{d(x, X \setminus \{x\}) \leq r | x \in X\}$ (rising starts later than Poisson)
- **F** CDF of the empty space distance: $F(r) = P\{d(u, X) \leq r\}$, for $u$ any grid crossing point
- **K** $\lambda K(r)$ is the mean number of points within a sphere of radius $r$ centered on each sample point: $K(r) = \text{Ripley’s function} = \frac{1}{\lambda} E[n(X \cap b(u, r) \setminus \{u\}) | u \in X]$, $\lambda$=intensity

E.g. ~ 20% of synapses have a NN at an inter-centroid distance of $\leq 500$ nm → some are side by side → the NN could reach neurotransmitter released and influence its behavior

**KS goodness-of-fit tests**

p-values $> 0.8$ for all samples when comparing observed vs. RSA

⇒ Spatial position of synaptic junctions is nearly random, only constrained by the fact that synapses cannot overlap

Perhaps nervous system highly ordered at other scales (macro, meso, microscopic) with specific targets (axon reaches specific cortical areas and layers); however random at the ultrastructural level (axon would form synapses randomly among its possible targets)
Spatial distribution of synapses

Asymmetric synapses (91% of the total found)

- Same conclusions when only asymmetric synapses are analyzed (symmetric are too few)
- Study whether asymmetric and symmetric synapses tend to form separate groups ⇔ their NN will tend to be of the same type (not intermingled at random)
- 2×2 contingency table:

| Type of synapse | Type of NN  
|-----------------|------------
| Asymmetric      | Observed   |
|                 | Expected   |
|                 |            |
| Asymmetric      | 704        |
|                 | 688.54     |
| Symmetric       | 43         |
|                 | 58.46      |

⇒ Fisher’s exact test rejects $H_0$ that both groups are intermingled at random (p<0.001, two tailed)

⇒ The number of symmetric synapses that have a symmetric nearest neighbor is 4.4 times higher than expected under $H_0$
Spatial distribution of synapses

Asymmetric synapses (91% of the total found)

- Same conclusions when only asymmetric synapses are analyzed (symmetric are too few)
- Study whether asymmetric and symmetric synapses tend to form separate groups ⇔ their NN will tend to be of the same type (not intermingled at random)

2 × 2 contingency table:

<table>
<thead>
<tr>
<th>Type of synapse</th>
<th>Type of NN</th>
<th>Asymmetric</th>
<th>Symmetric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymmetric</td>
<td>Observed</td>
<td>704</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>Expected</td>
<td>688.54</td>
<td>53.46</td>
</tr>
<tr>
<td>Symmetric</td>
<td>Observed</td>
<td>43</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Expected</td>
<td>58.46</td>
<td>4.54</td>
</tr>
</tbody>
</table>

⇒ Fisher’s exact test rejects $H_0$ that both groups are intermingled at random (p<0.001, two tailed)

⇒ The number of symmetric synapses that have a symmetric nearest neighbor is 4.4 times higher than expected under $H_0$
Outline

1. Introduction

2. Standard machine learning methods
   - Clustering for neuron classification
   - Clustering of spines
   - Knowledge discovery in Alzheimer’s disease
   - Supervised classification of neurons
   - Supervised classification of dementia development in Parkinson’s
   - Supervised classification for outcome prediction of epilepsy surgery

3. Advanced statistical and machine learning methods
   - EM based subspace clustering for discovering new types of neurons
   - Bayesian classifiers for probabilistic class neuron
   - Multi-dimensional classification for EQ-5D health states in Parkinson
   - Computer simulation of dendritic morphology with Bayesian networks
   - Spatial point processes for distribution of synapses
   - Local regularized regression to decode cognitive states in fMRI experiments

4. Optimization
   - Categorizing Parkinson disease based on non-motor symptoms
   - EDAs for parameter optimization in compartmental model
   - Multi-objective GAs for brain networks
   - Classifying MEG data with lasso logistic regression and EDAs

5. Conclusions

C. Bielza, P. Larrañaga

Machine Learning and Optimization in Neuroscience
fMRI experiments to decode cognitive states

- Functional Magnetic Resonance Imaging (fMRI) produces 3D images about brain activity (blood oxygen level) at a spatial resolution of a few mm.
- Star/Plus data collected at Carnegie Mellon U.: 6 people are shown in sequence a sentence and a picture. E.g. $\pm \ast$, “It is true that the plus is below the dollar”.
- Many brain areas involved: visual cortex for reading and seeing the sentence, Broca’s area for language processing, the Intra Parietal Sulcus for spatial visualization...
- Distinguish whether the subject is looking at a sentence or a picture during a particular time interval.
- 25 variables, the average activation of all voxels in each Region of Interest.
Lazy lasso for local regression

Diego Vidaurre · Concha Bielza · Pedro Larrañaga

Received: 17 February 2011 / Accepted: 13 July 2011
© Springer-Verlag 2011

Abstract Locally weighted regression is a technique that predicts the response for new data items from their neighbors in the training data set, where closer data items are assigned higher weights in the prediction. However, the original method may suffer from overfitting and fail to select the relevant variables. In this paper we propose combining a regularization approach with locally weighted regression to achieve sparse models. Specifically, the lasso is a shrinkage and selection method for linear regression. We present an algorithm that embeds lasso in an iterative procedure that alternatively computes weights and performs lasso-wise regression. The algorithm is tested on three synthetic scenarios and two real data sets. Results show that the proposed method outperforms linear and local models for several kinds of scenarios.
Locally weighted (linear) regression or loess

- Data set \( \mathcal{D} = \{(x^{(1)}, y^{(1)}), \ldots, (x^{(n)}, y^{(n)})\} \), with \( x^{(i)} = (x_1^{(i)}, x_2^{(i)}, \ldots, x_p^{(i)}) \)

- Response \( Y \) can’t always be predicted by means of a linear function of the covariates \( X_1, \ldots, X_p \Rightarrow \) local learning or nonparametric approach

- Locally weighted regression: for each \( x^{(l)} \) in the covariate space, \( \exists \) neighborhood where the regression surface is well approximated by a function

- For each \( x^{(l)} \), \( l \neq 1, \ldots, n \)

\[
\min_{\beta^{(l)}} \sum_{i=1}^{n} \left( w^{(l)}_i y^{(i)} - \sum_{j=1}^{p} w^{(l)}_i \beta^{(l)}_j x_j^{(i)} \right)^2
\]

- It’s a WRSS: \( w^{(l)}_i \) is (a function of) the distances of each data item \( x^{(i)} \) to \( x^{(l)} \). Weights attach more importance to closer points.

- \( w^{(l)}_i \) depends on a parameter (bandwidth) that controls how narrow/wide is the neighborhood. \( \beta_j= \) regression coefficients for \( x^{(l)} \)
Local regression and feature selection

FSS in loess

- Need to incorporate FSS into the loess methodology if required, i.e. with irrelevant variables (sparse influence of $X_1, \ldots, X_p$ on $Y$)

$\Rightarrow$ Combine loess and lasso ($l_1$-regularization)

$$\min_{\beta^{(l)}} \sum_{i=1}^{n} \left( w_i^{(l)} y^{(i)} - \sum_{j=1}^{p} w_i^{(l)} \beta_j^{(l)} x_j^{(i)} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j^{(l)}|$$

(1)

where $\lambda > 0$ is the regularization parameter that controls the amount of shrinkage (the larger, the greater the shrinkage and the smaller the $\beta_j$'s)

- Regularization prevents overfitting, reduces the variance of the estimates and gives more interpretable models. Moreover, lasso (least absolute shrinkage and selection operator) performs FSS

- However, distance calculation involves irrelevant variables $\Rightarrow$ leads to incorrect feature selections (incorrect weighting scheme) and inaccurate predictions
Local regression and feature selection

**Local prediction and local feature selection**

- Distances should only include relevant covariates: $1/0$, or perhaps in a smooth way, distances weighted by $\propto |\beta_j|$.
- ...but weights are computed prior to the regression, before knowing what variables are relevant for prediction!!
- **lazy lasso** proposal: algorithm that alternates variable selection and distance computation until some stopping criterion is met via no improvement in a (LOOCV) local version of prediction sum of squares or PRESS, commonly used for validation in local learning.

  - Compute distances with all variables and solve (1) with LARS algorithm [Efron et al., 2004]
  - Select the best vector of $\beta$’s among all $\lambda$’s according to some criterion (we use Mallows’ $C_p$ statistic)
  - **Update distances** weighted by $\propto |\beta_j|$. Weights are a function of distances: take $w_i = 1$ for the kNN of $x^{(l)}$ and $w_i = 0$ otherwise (to keep the local homoscedasticity assumption required). Solve (1)
  - ...
Results for Star/Plus experiment

- \( n = 40 \) trials \( \times 16 \) fMRI images = 640 data items per subject. 25 variables (ROIs)
- Consider nonlinear models (brain is complex) and sparse (which brain regions)
- \( Y \) is -1 ('picture') or 1 ('sentence') \( \Rightarrow \) Use regression for classification, based on the sign of the response. 0/1 error

<table>
<thead>
<tr>
<th>Subj.</th>
<th>LL</th>
<th>NLL</th>
<th>Error loess</th>
<th>lasso</th>
<th>RT</th>
<th>LL</th>
<th>NLL</th>
<th>lasso</th>
<th>RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.49</td>
<td>.5</td>
<td>.47</td>
<td>.43</td>
<td>.44</td>
<td>4(±6)</td>
<td>1.4(±2.7)*</td>
<td>19.8(±0.7)</td>
<td>18.9(±1.3)</td>
</tr>
<tr>
<td>2</td>
<td>.47</td>
<td>.5</td>
<td>.45</td>
<td>.44</td>
<td>.46</td>
<td>4.9(±5.7)</td>
<td>1.4(±2.3)*</td>
<td>9(±0.3)</td>
<td>18.3(±1)</td>
</tr>
<tr>
<td>3</td>
<td>.35</td>
<td>.45</td>
<td>.39</td>
<td>.4</td>
<td>.46</td>
<td>7.2(±7.5)</td>
<td>2.5(±4)*</td>
<td>24.9(±0.3)</td>
<td>17.5(±1.3)</td>
</tr>
<tr>
<td>4</td>
<td>.39</td>
<td>.46</td>
<td>.43</td>
<td>.38</td>
<td>.42</td>
<td>7.6(±7.2)</td>
<td>2.2(±3.9)*</td>
<td>11.4(±0.8)</td>
<td>19.2(±0.7)</td>
</tr>
<tr>
<td>5</td>
<td>.32</td>
<td>.39</td>
<td>.37</td>
<td>.35</td>
<td>.34</td>
<td>5.3(±4.4)</td>
<td>2.9(±3.1)*</td>
<td>22.7(±0.6)</td>
<td>15(±1.7)</td>
</tr>
<tr>
<td>6</td>
<td>.39</td>
<td>.52</td>
<td>.45</td>
<td>.4</td>
<td>.41</td>
<td>8(±7.2)</td>
<td>2.1(±3.7)*</td>
<td>15.4(±1.5)</td>
<td>19.8(±1.1)</td>
</tr>
</tbody>
</table>

LL= lazy lasso, NLL= solver of (1) with LARS, RT=regression tree

- Differences not statistically significant (from winner to the 2\( ^{nd} \) best) in error, but lasso and lazy lasso better than loess and RT
- NLL significantly better in sparseness (fewer ROIs), not in accuracy
- loess selects all variables
- With slightly worse accuracy vs. lasso, the lazy lasso exhibits a better FSS ability
Outline

1 Introduction

2 Standard machine learning methods
   - Clustering for neuron classification
   - Clustering of spines
   - Knowledge discovery in Alzheimer’s disease
   - Supervised classification of neurons
   - Supervised classification of dementia development in Parkinson’s
   - Supervised classification for outcome prediction of epilepsy surgery

3 Advanced statistical and machine learning methods
   - EM based subspace clustering for discovering new types of neurons
   - Bayesian classifiers for probabilistic class neuron
   - Multi-dimensional classification for EQ-5D health states in Parkinson
   - Computer simulation of dendritic morphology with Bayesian networks
   - Spatial point processes for distribution of synapses
   - Local regularized regression to decode cognitive states in fMRI experiments

4 Optimization
   - Categorizing Parkinson disease based on non-motor symptoms
   - EDAs for parameter optimization in compartmental model
   - Multi-objective GAs for brain networks
   - Classifying MEG data with lasso logistic regression and EDAs

5 Conclusions
Outline

1. Introduction

2. Standard machine learning methods
   - Clustering for neuron classification
   - Clustering of spines
   - Knowledge discovery in Alzheimer’s disease
   - Supervised classification of neurons
   - Supervised classification of dementia development in Parkinson’s
   - Supervised classification for outcome prediction of epilepsy surgery

3. Advanced statistical and machine learning methods
   - EM based subspace clustering for discovering new types of neurons
   - Bayesian classifiers for probabilistic class neuron
   - Multi-dimensional classification for EQ-5D health states in Parkinson
   - Computer simulation of dendritic morphology with Bayesian networks
   - Spatial point processes for distribution of synapses
   - Local regularized regression to decode cognitive states in fMRI experiments

4. Optimization
   - Categorizing Parkinson disease based on non-motor symptoms
   - EDAs for parameter optimization in compartmental model
   - Multi-objective GAs for brain networks
   - Classifying MEG data with lasso logistic regression and EDAs

5. Conclusions
Motor and non-motor symptoms

- Loss of dopaminergic neurons
- **Dopamine shortage** causes the classical symptoms: trembling and slow physical movements
- Exact cause is still unknown
- J. Parkinson (1817)

**Relationship between motor and non-motor symptoms**

- Well established severity scale based on motor symptoms [Hoehn and Yahr, 2004])
- Non motor symptoms start before motor symptoms
- Objective: predict Parkinson disease stage (following Hoehn and Yahr severity scale) based on non-motor symptoms
Motor and non-motor severity scales

### Hoehn-Yahr scale

1. Unilateral involvement only usually with minimal or no functional disability
2. Bilateral involvement without impairment of balance
3. Bilateral disease: mild to moderate disability with impaired postural reflexes; physically independent
4. Severely disabling disease; still able to walk or stand unassisted
5. Confinement to bed or wheelchair unless aided

### Other non-motor scales (additives)

- Non-motor symptoms (NMS). 9 items: $\in \{0, \ldots, 360\}$ - symptom’s frequency $\times$ severity degree
- Psychiatric complications (SCOPA-PC). 7 items: $\in \{0, \ldots, 21\}$
- Autonomous body functions (SCOPA-AUT). 25 items: $\in \{0, \ldots, 71\}$
- SCOPA cognitive (SCOPA-COG). 10 items: $\in \{0, \ldots, 43\}$ – the lower value ⇒ the worse condition

In preparation
Optimization problem - Find \( k = (k_1, k_2, k_3, k_4) \)

<table>
<thead>
<tr>
<th>HOEHN-YAHР</th>
<th>([0,k_1])</th>
<th>([k_1+1,k_2])</th>
<th>([k_2+1,k_3])</th>
<th>([k_3+1,k_4])</th>
<th>([k_4+1,\text{max}])</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(c_{11})</td>
<td>(c_{12})</td>
<td>(c_{13})</td>
<td>(c_{14})</td>
<td>(c_{15})</td>
</tr>
<tr>
<td>2</td>
<td>(c_{21})</td>
<td>(c_{22})</td>
<td>(c_{23})</td>
<td>(c_{24})</td>
<td>(c_{25})</td>
</tr>
<tr>
<td>3</td>
<td>(c_{31})</td>
<td>(c_{32})</td>
<td>(c_{33})</td>
<td>(c_{34})</td>
<td>(c_{35})</td>
</tr>
<tr>
<td>4</td>
<td>(c_{41})</td>
<td>(c_{42})</td>
<td>(c_{43})</td>
<td>(c_{44})</td>
<td>(c_{45})</td>
</tr>
<tr>
<td>5</td>
<td>(c_{51})</td>
<td>(c_{52})</td>
<td>(c_{53})</td>
<td>(c_{54})</td>
<td>(c_{55})</td>
</tr>
</tbody>
</table>

For \textbf{NMS, SCOPA-PC, SCOPA-AUT}: \( k^* = \arg \max_k \sum_{i=1}^{5} c_{ii} \)

For \textbf{SCOPA-COG}: \( k^* = \arg \max_k \sum_{i=1}^{5} c_{i(6-i)} \)
Exhaustive search of the cut-points

411 patients

<table>
<thead>
<tr>
<th>HOEHN-YAHR</th>
<th>NMS $k^*$ = (5, 57, 141, 198) with 45% acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[0,5]</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HOEHN-YAHR</th>
<th>SCOPA-COG $k^*$ = (3, 10, 13, 39) with 46% acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[0,3]</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>
Indices subset selection followed by cut points (exhaustive searches)

**NMS (\(512 = 2^9\) indices subset selection)**

<table>
<thead>
<tr>
<th>Mild (1 + 2)</th>
<th>[0,43]</th>
<th>[44,57]</th>
<th>[58,144]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>216</td>
<td>7</td>
<td>5</td>
</tr>
</tbody>
</table>

**Hoehn-Yahr**

<table>
<thead>
<tr>
<th>Moderate (3)</th>
<th>[0,43]</th>
<th>[44,57]</th>
<th>[58,144]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>107</td>
<td>21</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Severe (4 + 5)</th>
<th>[0,43]</th>
<th>[44,57]</th>
<th>[58,144]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>28</td>
<td>8</td>
<td>14</td>
</tr>
</tbody>
</table>

**SCOPA-COG (1,024 = 2^{10} indices subset selection)**

<table>
<thead>
<tr>
<th>[0,2]</th>
<th>[3,3]</th>
<th>[4,8]</th>
<th>[9,16]</th>
<th>[17,17]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>25</td>
<td>131</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>43</td>
<td>66</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>6</td>
<td>12</td>
<td>23</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>
Goal: Predict Hoehn-Yahr (mild, moderate, severe) using knowledge from different domains of the non-motor indexes

New data matrix: 411 Parkinson patients + 87 predictor variables and 1 class variable

Avoiding linearity

1. Classification paradigms: Naïve Bayes (NB), k-nearest neighbor (k-NN), linear discriminant analysis (LDA), C4.5

2. Feature subset selection: Wrapper search using an estimation of distribution algorithm (EDA-UMDA)

3. Validation scheme: Score guided by internal 10-fold cross-validation
Metaheuristic optimization

- Evolutionary Computation
  - Genetic programming
  - Genetic algorithms
  - Evolution strategies
  - Differential evolution
  - Scatter search
- Particle swarm optimization
- Ant colony optimization
- Estimation of distribution algorithms
- Simulated annealing
- GRASP
- Iterated local search
- Stochastic local search
- Guided local search

Dynamic objective function

C. Bielza, P. Larrañaga
Machine Learning and Optimization in Neuroscience
Introduction to EDAs and multi-objective optimization

Estimation of distribution algorithms (EDAs). Larrañaga and Lozano (2001)

- **Initial population of candidate solutions**
- **Selected candidates**
- **Bayesian or Gaussian network learning**
- **Sampling of new candidate solutions**

C. Bielza, P. Larrañaga

Machine Learning and Optimization in Neuroscience
Introduction to EDAs and multi-objective optimization

Estimation of distribution algorithms (EDAs). Larrañaga and Lozano (2001)

Initial population of candidate solutions

Selected candidates

Bayesian or Gaussian network learning

Sampling of new candidate solutions
Introduction to EDAs and multi-objective optimization


- **Probabilistic model:** $p_I(x) = \prod_{i=1}^{n} p_I(x_i)$
- **Structural learning:** not necessary

**Bivariate EDAs: Mutual Information Maximization for Input Clustering (MIMIC). De Bonet et al. (1997)**

- **Probabilistic model:**
  
  $$
  p_I^\pi(x) = p_I(x_{i_1} \mid x_{i_2})p_I(x_{i_2} \mid x_{i_3}) \cdot \cdot \cdot p_I(x_{i_{n-1}} \mid x_{i_n})p_I(x_{i_n})
  $$

- **Structural learning:** best permutation (factorization closest to the empirical distribution in the sense of Kullback-Leibler divergence)

**Multivariate EDAs: Estimation of Bayesian Network Algorithm (EBNA). Etxeberria and Larrañaga (1999)**

- **Probabilistic model:** $p_I(x) = \prod_{i=1}^{n} p_I(x_i \mid pa_i)$
- **Structural learning:** directed acyclic graph
### Some results

#### mild VS. moderate VS. severe

<table>
<thead>
<tr>
<th></th>
<th>NB</th>
<th>k-NN</th>
<th>LDA</th>
<th>C4.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>69.81%</td>
<td>72.50%</td>
<td>69.00%</td>
<td>69.00%</td>
</tr>
<tr>
<td>Subset size</td>
<td>18</td>
<td>17</td>
<td>14</td>
<td>3</td>
</tr>
</tbody>
</table>

#### mild VS. moderate

<table>
<thead>
<tr>
<th></th>
<th>NB</th>
<th>k-NN</th>
<th>LDA</th>
<th>C4.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>78.76%</td>
<td>78.76%</td>
<td>77.53%</td>
<td>76.00%</td>
</tr>
<tr>
<td>Subset size</td>
<td>10</td>
<td>15</td>
<td>10</td>
<td>6</td>
</tr>
</tbody>
</table>

#### moderate VS. severe

<table>
<thead>
<tr>
<th></th>
<th>NB</th>
<th>k-NN</th>
<th>LDA</th>
<th>C4.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>81.01%</td>
<td>86.07%</td>
<td>79.74%</td>
<td>86.07%</td>
</tr>
<tr>
<td>Subset size</td>
<td>9</td>
<td>16</td>
<td>9</td>
<td>6</td>
</tr>
</tbody>
</table>
Important features

Things are getting worse in PD when ...

- **Scpc1** - Hallucinations
- **Scpc2** - Illusions and misidentification of persons
- **Scaulcho** - Have you had difficulty swallowing or have you chocked?
- **Nms25** - Does the patient have altered interest in sex?
Outline

1. Introduction

2. Standard machine learning methods
   - Clustering for neuron classification
   - Clustering of spines
   - Knowledge discovery in Alzheimer’s disease
   - Supervised classification of neurons
   - Supervised classification of dementia development in Parkinson’s
   - Supervised classification for outcome prediction of epilepsy surgery

3. Advanced statistical and machine learning methods
   - EM based subspace clustering for discovering new types of neurons
   - Bayesian classifiers for probabilistic class neuron
   - Multi-dimensional classification for EQ-5D health states in Parkinson
   - Computer simulation of dendritic morphology with Bayesian networks
   - Spatial point processes for distribution of synapses
   - Local regularized regression to decode cognitive states in fMRI experiments

4. Optimization
   - Categorizing Parkinson disease based on non-motor symptoms
   - EDAs for parameter optimization in compartmental model
   - Multi-objective GAs for brain networks
   - Classifying MEG data with lasso logistic regression and EDAs

5. Conclusions
Single compartmental models

- **Compartmental models** [Hodgkin and Huxley, 1952], based on a circuit representation of a neuron, describe the potential along the dendritic tree.

- The model consists of a system of ordinary differential equations.

- The model is used to replicate the response of a real neuron to a set of simple current stimuli.

- **Single compartmental models** neglects the neuron’s spatial structure and focuses on how its various ionic currents contribute to its behavior.
Neuronal parameter optimization

The optimization problem in compartmental models

- The process of identifying sets of parameters that lead to a desired electrical activity pattern in a neuron or neural network model that is not fully constrained by experimental data [Printz, 2007]

Related questions

- Choice of an objective function or goodness measure to evaluate the possible solutions
- Selection of an appropriate optimization algorithm
- Analyze the interactions between the parameters
Neuronal parameter optimization

Objective functions

- **Point-to-point comparison of voltage traces:** comparison of the electrophysiological traces (data vs. model)

- **Feature-based:** (1) spike rate; (2) an accommodation index; (3) latency to first spike; (4) average AP overshoot; (5) average depth of after hyperpolarization (AHP); (6) average AP width

- **Multi-objective:** several feature-based criteria at the same time

Hybrid method [van Geit et al., 2007]
Neuronal parameter optimization

Conductance-based compartmental neuron model

Data obtained from lobster stomatogastric neurons [Prinz et al., 2003]

Neuron potential $V$ is modeled by differential equations, based on each of the model membrane currents and 8 input currents:

- $I_{Na}$: $Na^+$ current
- $I_{CaT}$ and $I_{CaS}$: two $Ca^{2+}$ currents
- $I_{A}$: a transient $K^+$ current
- $I_{KCa}$: a $Ca^{2+}$ dependent $K^+$ current
- $I_{Kd}$: a delayed rectifier $K^+$ current
- $I_{H}$: a hyperpolarization-activated inward current
- $I_{leak}$: a leak current

Database of $1,679,616 = 6^8$ neuron models

<table>
<thead>
<tr>
<th>Var</th>
<th>Current</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>$I_{Na}$</td>
<td>0</td>
<td>100.00</td>
<td>200.00</td>
<td>300.00</td>
<td>400.00</td>
<td>500.00</td>
</tr>
<tr>
<td>$X_2$</td>
<td>$I_{CaT}$</td>
<td>0</td>
<td>2.50</td>
<td>5.00</td>
<td>7.50</td>
<td>10.00</td>
<td>12.50</td>
</tr>
<tr>
<td>$X_3$</td>
<td>$I_{CaS}$</td>
<td>0</td>
<td>2.00</td>
<td>4.00</td>
<td>6.00</td>
<td>8.00</td>
<td>10.00</td>
</tr>
<tr>
<td>$X_4$</td>
<td>$I_{A}$</td>
<td>0</td>
<td>10.00</td>
<td>20.00</td>
<td>30.00</td>
<td>40.00</td>
<td>50.00</td>
</tr>
<tr>
<td>$X_5$</td>
<td>$I_{KCa}$</td>
<td>0</td>
<td>5.00</td>
<td>10.00</td>
<td>15.00</td>
<td>20.00</td>
<td>25.00</td>
</tr>
<tr>
<td>$X_6$</td>
<td>$I_{Kd}$</td>
<td>0</td>
<td>25.00</td>
<td>50.00</td>
<td>75.00</td>
<td>100.00</td>
<td>125.00</td>
</tr>
<tr>
<td>$X_7$</td>
<td>$I_{H}$</td>
<td>0</td>
<td>0.01</td>
<td>0.02</td>
<td>0.03</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>$X_8$</td>
<td>$I_{leak}$</td>
<td>0</td>
<td>0.01</td>
<td>0.02</td>
<td>0.03</td>
<td>0.04</td>
<td>0.05</td>
</tr>
</tbody>
</table>
SANTANA R., BIELZA C., LARRAÑAGA P. (2011)
Using probabilistic dependencies improves the search of conductance-based compartmental neuron models
*Lecture Notes in Computer Science 6023, EvoBIO 2010*, pp. 170-181, Springer
Objective function and metaheuristic

- Given a target neuron, our goal is to find, from a large set of candidates (6^8 neuron models), a neuron model that resembles its electrical activity as described by recorded experimental data.

- Let $x^*$ and $x$ respectively be the target neuron and any other neuron of the search space. The fitness function $f(x)$:

$$f(x) = (fq_s(x) -fq_s(x^*))^2 + (fq_3(x) -fq_3(x^*))^2 + (fq_5(x) -fq_5(x^*))^2$$

  - $fq_s$: the spontaneous frequency
  - $fq_3$: the steady state maximal frequencies as computed during 3nA current injection
  - $fq_5$: the steady state maximal frequencies as computed during 5nA current injection

- Metaheuristic: three types of estimation of distribution algorithms (EDAs) (UMDA, TREE-EDA, and EBNA)
Neuronal parameter optimization

Optimization results

- We selected 10 (target) neurons from the large data set.
- For each neuron: 30 experiments of UMDA, TREE-EDA and EBNA.

<table>
<thead>
<tr>
<th></th>
<th>$I_{Na}$</th>
<th>$I_{CaT}$</th>
<th>$I_{CaS}$</th>
<th>$I_A$</th>
<th>$I_{KCa}$</th>
<th>$I_{Kd}$</th>
<th>$I_H$</th>
<th>$I_{leak}$</th>
<th>$f_{qs}(x^*)$</th>
<th>$f_{q3}(x^*)$</th>
<th>$f_{q6}(x^*)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>100</td>
<td>7.5</td>
<td>4</td>
<td>40</td>
<td>5</td>
<td>125</td>
<td>0.00</td>
<td>0.01</td>
<td>1.1632</td>
<td>54.6364</td>
<td>60.1949</td>
</tr>
<tr>
<td>B</td>
<td>500</td>
<td>5.0</td>
<td>6</td>
<td>40</td>
<td>10</td>
<td>125</td>
<td>0.00</td>
<td>0.01</td>
<td>1.1559</td>
<td>37.7609</td>
<td>42.2354</td>
</tr>
<tr>
<td>C</td>
<td>100</td>
<td>12.5</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>25</td>
<td>0.04</td>
<td>0.01</td>
<td>215.0538</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>300</td>
<td>5.0</td>
<td>8</td>
<td>0</td>
<td>25</td>
<td>0</td>
<td>0.05</td>
<td>0.04</td>
<td>4.5382</td>
<td>5.8246</td>
<td>7.8225</td>
</tr>
<tr>
<td>E</td>
<td>100</td>
<td>7.5</td>
<td>4</td>
<td>10</td>
<td>0</td>
<td>25</td>
<td>0.05</td>
<td>0.03</td>
<td>68.2173</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F</td>
<td>300</td>
<td>10.0</td>
<td>10</td>
<td>40</td>
<td>20</td>
<td>25</td>
<td>0.02</td>
<td>0.03</td>
<td>3.9518</td>
<td>23.1826</td>
<td>27.9265</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>2.5</td>
<td>8</td>
<td>40</td>
<td>5</td>
<td>100</td>
<td>0.02</td>
<td>0.01</td>
<td>2.4422</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>H</td>
<td>300</td>
<td>0.0</td>
<td>10</td>
<td>20</td>
<td>15</td>
<td>100</td>
<td>0.05</td>
<td>0.01</td>
<td>18.0810</td>
<td>37.0142</td>
<td>42.6758</td>
</tr>
<tr>
<td>I</td>
<td>100</td>
<td>0.0</td>
<td>4</td>
<td>30</td>
<td>25</td>
<td>75</td>
<td>0.00</td>
<td>0.02</td>
<td>4.8984</td>
<td>26.9808</td>
<td>35.5637</td>
</tr>
<tr>
<td>J</td>
<td>400</td>
<td>12.5</td>
<td>4</td>
<td>40</td>
<td>20</td>
<td>125</td>
<td>0.00</td>
<td>0.04</td>
<td>8.0523</td>
<td>35.9712</td>
<td>47.1328</td>
</tr>
</tbody>
</table>

Setting I (top) and setting II (bottom)

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>UMDA</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>TREE-EDA</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>EBNA</td>
<td>0</td>
<td>1</td>
<td>30</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>18</td>
<td>62</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>UMDA</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>TREE-EDA</td>
<td>0</td>
<td>0</td>
<td>19</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>23</td>
</tr>
<tr>
<td>EBNA</td>
<td>2</td>
<td>0</td>
<td>30</td>
<td>4</td>
<td>12</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>25</td>
<td>83</td>
</tr>
</tbody>
</table>
Neuronal parameter optimization

Structures of the interactions captured by EBNA

a) Most frequent structures for instance 1071411
b) Most frequent for at least 5 of the 10
Outline

1. Introduction

2. Standard machine learning methods
   - Clustering for neuron classification
   - Clustering of spines
   - Knowledge discovery in Alzheimer’s disease
   - Supervised classification of neurons
   - Supervised classification of dementia development in Parkinson’s
   - Supervised classification for outcome prediction of epilepsy surgery

3. Advanced statistical and machine learning methods
   - EM based subspace clustering for discovering new types of neurons
   - Bayesian classifiers for probabilistic class neuron
   - Multi-dimensional classification for EQ-5D health states in Parkinson
   - Computer simulation of dendritic morphology with Bayesian networks
   - Spatial point processes for distribution of synapses
   - Local regularized regression to decode cognitive states in fMRI experiments

4. Optimization
   - Categorizing Parkinson disease based on non-motor symptoms
   - EDAs for parameter optimization in compartmental model
   - Multi-objective GAs for brain networks
   - Classifying MEG data with lasso logistic regression and EDAs

5. Conclusions
Extraction of a real brain structural connectivity network

MRI Acquisition

1. Diffusion Spectrum Imaging
2. T1w high res.
3a. Segmentation
3b. Partition into 66 anatomical subregions
4. Tractography
5. Partition into 1000 ROIs

Whole brain structural connection network
Brain activity can be modeled as a dynamical process acting on a network. Each vertex of the structure represents an elementary component (brain areas, groups of neurons or individual cells).

The network topology is a key element for understanding the behavior of the process.

Based on real brain networks, similar artificial brain networks can be obtained via optimization. The artificial brain networks should satisfy some constraints identified in the real network. The similarity is measured in terms of predefined topological characteristics (structural motifs and functional motifs).
A (structural) “motif” is a connected graph or network consisting of $M$ vertices and a set of edges (arcs) (maximally $M^2 - M$ for directed graphs, minimally $M - 1$) forming a subgraph of a larger network. The name structural is because a larger network could be structurally assembled from a finite set of such structural motifs.

A functional motif of a given structural motif consists of the original $M$ vertices of the structural motif, but contains only a subset of its edges.
Optimization approach

- **Structural optimization** in brain networks implies the identification of a network topology that:
  - satisfies a number of **constraints** identified in the real brain network (same number of in and out node degrees)
  - is **optimal** with respect to **measure(s)** defined in the space of networks (structural motifs and functional motifs)

- **Multiobjective optimization** under constraints
  - **Maximize** the number of functional motifs (**functionally** of the brain)
  - **Minimize** the number of structural motifs (**simplicity** of the brain)
Optimizing Brain Networks Topologies Using Multi-objective Evolutionary Computation

Roberto Santana · Concha Bielza · Pedro Larrañaga

Published online: 30 September 2010
© Springer Science+Business Media, LLC 2010

Abstract The analysis of brain network topological features has served to better understand these networks and reveal particular characteristics of their functional behavior. The distribution of brain network motifs is particularly useful for detecting and describing differences between brain networks

Keywords Brain networks · Evolutionary algorithm · Network motifs · Multi-objective optimization · Network optimization
Structural optimization in brain networks

Optimizing brain networks topologies using multi-objective GAs

- A multi-objective evolutionary optimization approach to generate optimized artificial networks
- The Pareto set approximation is used
- Original brain networks have a reduced structural motif number and a high functional motif number, but they are not optimal
- Genetic algorithm (with a rewiring operator ([Sporns and Kötter, 2004] that guarantees the constraints) + local optimizer + Pareto-ranking selection

Multi-objective optimization. Pareto front

C. Bielza, P. Larrañaga
Machine Learning and Optimization in Neuroscience
Structural optimization in brain networks

Optimizing brain networks topologies using multi-objective GAs

| Original brain network | Cortex areas | \( n \) | \(|E|\) |
|------------------------|-------------|--------|------|
| macaque30              | visual      | 30     | 311  |
| macaque32              | visual      | 32     | 315  |
| macaque47              | visual and sensorimotor | 47     | 505  |
| macaque71              | visual and sensorimotor | 71     | 746  |

- **Population size:** 500 individuals
- **Number of generations:** 200 for macaque30, 200 for macaque32, 500 for macaque47, 1000 for macaque71
- **Number of runs:** 20

C. Bielza, P. Larraña
Machine Learning and Optimization in Neuroscience
Structural optimization in brain networks

Optimizing brain networks topologies using multi-objective GAs

Structural motif number against functional motif number for 10,000 randomly sampled networks (red circles) and original brain networks (blue triangle)

C. Bielza, P. Larrañaga
Structural optimization in brain networks

Optimizing brain networks topologies using multi-objective GAs

Structural motif number against functional motif number for non-dominated solutions learned in each of the 20 runs of the GA (blue), and absolute non-dominated solutions (red) and original brain networks (red triangle)
Outline

1. Introduction
2. Standard machine learning methods
   - Clustering for neuron classification
   - Clustering of spines
   - Knowledge discovery in Alzheimer’s disease
   - Supervised classification of neurons
   - Supervised classification of dementia development in Parkinson’s
   - Supervised classification for outcome prediction of epilepsy surgery
3. Advanced statistical and machine learning methods
   - EM based subspace clustering for discovering new types of neurons
   - Bayesian classifiers for probabilistic class neuron
   - Multi-dimensional classification for EQ-5D health states in Parkinson
   - Computer simulation of dendritic morphology with Bayesian networks
   - Spatial point processes for distribution of synapses
   - Local regularized regression to decode cognitive states in fMRI experiments
4. Optimization
   - Categorizing Parkinson disease based on non-motor symptoms
   - EDAs for parameter optimization in compartmental model
   - Multi-objective GAs for brain networks
   - Classifying MEG data with lasso logistic regression and EDAs
5. Conclusions
Classifying MEG data

- Magnetoencephalography (MEG) records magnetic fields produced by electrical currents occurring naturally in the brain.
- Magnetic fields are less distorted than electric fields by the skull and scalp ⇒ better spatial resolution of MEG than EEG.
- MEG is more sensitive to superficial cortical activity ⇒ useful for the study of neocortical epilepsy.
- Typically for controlling a BCI (translate electrical signals into commands without motor intervention).
Classifying MEG data. The problem

MEG data analysis competition: challenge

- Challenge at **BIOMAG-2010**: Apply a classifier to predict in which direction (right-left) 4 subjects were covertly attending (at the single trial level)
- **1-D version** (right-left) of the more general problem with four directions: top, right, bottom, left
- No **prior knowledge** about the physiological mechanisms that determine the brain activity is used
- Not only **max accuracy** of the mental task prediction but also investigate how information from different brain areas **contributes to the predictions**
- **Subject and trial variability** are frequently a source of poor BCI performance
Classifying MEG data. The problem

MEG data analysis competition: data [van Gerven et al., 2009]

- **128 trials** collected per condition (right-left) for each of the 4 subjects
- **Each trial**: a 400 ms presentation of the cue, then subjects had 2.5 s to covertly attend (i.e., without moving their eyes from the fixation cross) in the indicated direction; the corresponding square turned either green or red for 40 ms

- A **rest period** of 1.5 s between trials
- A time series output from each of 256 trials (128 × 2), 4 subjects, 274 MEG channels (total: **280,576 time series** each with 600-component numerical vectors)
- Rules: report the **classification rate** (correctly classified trials) for each subject, computed using leave-one-out cross-validation and report the classification procedure
SANTANA R., BIELZA C., LARRAÑAGA P. (2011)
Regularized logistic regression and
multi-objective variable selection for classifying
MEG data
International Journal of Human-Computer
Studies
submitted
Classifying MEG data: our approach

Four different types of information for classification

- **Raw** data (averaging a 10 dimensional window): from 600 time points to 60 by averaging a time window comprising 10 points $\Rightarrow 60 \times 274 = 16,440$ features for the classifier.

- Channel correlation: **Correlation matrix** between the time series corresponding to all the channels for a given trial $\Rightarrow \frac{274 \times 273}{2} = 37,401$ features (upper triangular part of the correlation matrix).

- **Topological measures of the interaction graph** (strong correlations: $> 0.5$ or $< -0.5$) to reveal local and global information not directly recognizable from the correlation values: (average pair distance, mean indegree of the vertices, clustering coefficients, ...) $\Rightarrow 274 \times 13 + 7 = 3,569$ features.

- **Combined representation**: Raw data and channel correlations $\Rightarrow 16,440 + 37,401 = 53,841$ features.
Classifying MEG data: our approach

Lasso regularized logistic regression

- Let $\pi_j$ denote $P(C = 1|x_j), j = 1, \ldots, N$. The logistic regression model is defined

$$\log \frac{\pi_j}{1 - \pi_j} = \beta_0 + \sum_{i=1}^{p} \beta_i x_{ji}$$

- The log-likelihood function is: $\log \mathcal{L}(\beta) = \sum_{j=1}^{N} (c_j \log \pi_j + (1 - c_j) \log(1 - \pi_j))$
where $c_j$ denotes the response values

- Lasso-type regularized logistic regression, able to deal with thousands of features

$$\max_{\beta} \left[ \log \mathcal{L}(\beta) - \lambda \sum_{j=1}^{p} |\beta_j| \right]$$

- $\lambda$ selection: $\lambda$ that maximizes the training set accuracy
Classifying MEG data: our approach

Channels from which information is extracted

- Two scenarios: **all channels** (274) vs. **occipitoparietal** channels (86)
- Channel set selection problem as a way to improve classification results
Classifying MEG data: our approach

Selection of optimal channels

- Simultaneously optimize the accuracy for the 4 subjects with multi-objective optimization
- **Encoding**: A binary vector \( \mathbf{x} = (x_1, \ldots, x_{274}) \) represents a possible subset of channels (\( x_i = 1 \) means channel \( i \) has been selected)
- **Pareto set** of solutions (non-dominated)
- **Evolutionary computation** (Pareto ranking selection):
  - Genetic algorithm: one-point crossover + bit mutation
  - EDAs: UMDA and Tree
Classifying MEG data: our approach

One of the 152 solutions in the Pareto set

<table>
<thead>
<tr>
<th>Subject</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.8701</td>
<td>0.8228</td>
<td>0.7638</td>
<td>0.7362</td>
</tr>
</tbody>
</table>

Analysis of the Pareto set solutions

- **Scalp locations** of channels and their relative **frequency** in the Pareto set solutions
- Most around the **occipitoparietal** region, but also e.g. the **frontal** area
Outline

1. Introduction
2. Standard machine learning methods
   - Clustering for neuron classification
   - Clustering of spines
   - Knowledge discovery in Alzheimer's disease
   - Supervised classification of neurons
   - Supervised classification of dementia development in Parkinson's
   - Supervised classification for outcome prediction of epilepsy surgery
3. Advanced statistical and machine learning methods
   - EM based subspace clustering for discovering new types of neurons
   - Bayesian classifiers for probabilistic class neuron
   - Multi-dimensional classification for EQ-5D health states in Parkinson
   - Computer simulation of dendritic morphology with Bayesian networks
   - Spatial point processes for distribution of synapses
   - Local regularized regression to decode cognitive states in fMRI experiments
4. Optimization
   - Categorizing Parkinson disease based on non-motor symptoms
   - EDAs for parameter optimization in compartmental model
   - Multi-objective GAs for brain networks
   - Classifying MEG data with lasso logistic regression and EDAs
5. Conclusions
Challenging problems (standard and new) in modeling and optimization

- **Brain** as the most complex organ
- **Machine learning**
  - CLUSTERING: affinity propagation, hierarchical
  - SUPERVISED CLASSIFICATION: naive Bayes, classification trees, $k$-NN, multi-layer perceptron, lasso logistic regression, SVMs
  - FEATURE SUBSET SELECTION: filter (CFS), wrapper (GAs)
  - SEMI-SUPERVISED WITH CLASS DISCOVERY: subspace EM
  - SUPERVISED CLASSIFICATION WITH PROBABILISTIC LABELS: naive Bayes, AODE
  - MULTI-DIMENSIONAL CLASSIFICATION: multi-dimensional Bayesian classifiers
  -JOINT PROBABILITY DISTRIBUTION: Bayesian networks
  -SPATIAL POINT PROCESSES: complete spatial randomness (CSR), random sequential adsorption (RSA)
  -LOCALLY WEIGHTED LINEAR REGRESSION

- **Optimization**
  - GENETIC ALGORITHMS
  - ESTIMATION OF DISTRIBUTION ALGORITHMS
  -MULTIOBJECTIVE OPTIMIZATION WITH EVOLUTIONARY COMPUTATION
Thanks to Cajal Blue Brain members

Universidad Politécnica de Madrid

Instituto Cajal
Many thanks to

1 Colleagues:
   - L. Alonso-Nanclares
   - R. Armañanzas
   - F. Baguear
   - L. Baumela
   - R. Benavides-Piccione
   - H. Borchani
   - J. DeFelipe
   - B. Gradl
   - L. Guerra
   - P. López-Cruz
   - P. Martínez
   - L. McGarry
   - A. Merchán
   - D. Morales
   - V. Robles
   - R. Santana
   - R. Yuste

2 Projects:
   - Cajal Blue Brain (Spanish Ministry of Science and Innovation)
   - Consolider Ingenio 2010-CSD2007-00018
   - TIN2010-20900-C04-04 (Spanish Ministry of Science and Innovation)
MACHINE LEARNING AND OPTIMIZATION IN NEUROSCIENCE

Concha Bielza, Pedro Larrañaga

Computational Intelligence Group
Artificial Intelligence Department
Technical University of Madrid, Spain

CAEPIA’11
Tenerife, 7 de noviembre de 2011