BAYESIAN NETWORKS BASED
MULTI-DIMENSIONAL CLASSIFICATION

Pedro Larrañaga

Computational Intelligence Group
Artificial Intelligence Department
Technical University of Madrid

3rd International Conference on Big Data Analytics, Data Mining and Computational Intelligence
Madrid, July 18, 2018
Outline

1. Multi-label classification and multi-dimensional classification
2. Binary relevance and label power set
3. Bayesian networks
4. Multi-dimensional Bayesian network classifiers
5. Applications
6. Conclusions
7. References
Outline

1. Multi-label classification and multi-dimensional classification
2. Binary relevance and label power set
3. Bayesian networks
4. Multi-dimensional Bayesian network classifiers
5. Applications
6. Conclusions
7. References
Simultaneous object recognition in images (multi-label)
Medical diagnosis (multi-label)

WITH RESPECT, MY MEDICAL DEGREE TRUMPS YOUR GOOGLE-DIAGNOSIS SEARCH
Multiple fault diagnosis (multi-label)
Single label classification versus multi-label classification

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
<th>$C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2</td>
<td>1.4</td>
<td>4.7</td>
<td>7.5</td>
<td>3.7</td>
<td>1</td>
</tr>
<tr>
<td>2.8</td>
<td>6.3</td>
<td>1.6</td>
<td>4.7</td>
<td>2.7</td>
<td>0</td>
</tr>
<tr>
<td>7.7</td>
<td>6.2</td>
<td>4.1</td>
<td>3.3</td>
<td>7.7</td>
<td>1</td>
</tr>
<tr>
<td>9.2</td>
<td>0.4</td>
<td>2.8</td>
<td>0.5</td>
<td>3.9</td>
<td>0</td>
</tr>
<tr>
<td>5.5</td>
<td>5.3</td>
<td>4.9</td>
<td>0.6</td>
<td>6.6</td>
<td>1</td>
</tr>
</tbody>
</table>

Single label classification

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
<th>$C_1$</th>
<th>$C_2$</th>
<th>$C_3$</th>
<th>$C_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2</td>
<td>1.4</td>
<td>4.7</td>
<td>7.5</td>
<td>3.7</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2.8</td>
<td>6.3</td>
<td>1.6</td>
<td>4.7</td>
<td>2.7</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7.7</td>
<td>6.2</td>
<td>4.1</td>
<td>3.3</td>
<td>7.7</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>9.2</td>
<td>0.4</td>
<td>2.8</td>
<td>0.5</td>
<td>3.9</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5.5</td>
<td>5.3</td>
<td>4.9</td>
<td>0.6</td>
<td>6.6</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Multi-label classification
Multi-label classification vs multi-dimensional classification

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
<th>$C_1$</th>
<th>$C_2$</th>
<th>$C_3$</th>
<th>$C_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2</td>
<td>1.4</td>
<td>4.7</td>
<td>7.5</td>
<td>3.7</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2.8</td>
<td>6.3</td>
<td>1.6</td>
<td>4.7</td>
<td>2.7</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7.7</td>
<td>6.2</td>
<td>4.1</td>
<td>3.3</td>
<td>7.7</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>9.2</td>
<td>0.4</td>
<td>2.8</td>
<td>0.5</td>
<td>3.9</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5.5</td>
<td>5.3</td>
<td>4.9</td>
<td>0.6</td>
<td>6.6</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Multi-label classification

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
<th>$C_1$</th>
<th>$C_2$</th>
<th>$C_3$</th>
<th>$C_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2</td>
<td>1.4</td>
<td>4.7</td>
<td>7.5</td>
<td>3.7</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>2.8</td>
<td>6.3</td>
<td>1.6</td>
<td>4.7</td>
<td>2.7</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7.7</td>
<td>6.2</td>
<td>4.1</td>
<td>3.3</td>
<td>7.7</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>9.2</td>
<td>0.4</td>
<td>2.8</td>
<td>0.5</td>
<td>3.9</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>5.5</td>
<td>5.3</td>
<td>4.9</td>
<td>0.6</td>
<td>6.6</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

Multi-dimensional classification
Neuronal cell-type classification (multi-dimensional)

- $X_1, \ldots, X_n$ are morphological variables
- $C_1, \ldots, C_d$
  - Animal specie: rat, mouse, monkey, cat, human, ...
  - Neuronal cell-type: pyramidal, interneuron, Purkinje, Martinotti,...
  - Brain region: amygdala, cerebral cortex, hippocampus, ...
  - Age of the animal: young, adult
A multi-label (multi-dimensional) data set
\[ \mathcal{D} = \{(\mathbf{x}^{(1)}, \mathbf{c}^{(1)}), \ldots, (\mathbf{x}^{(N)}, \mathbf{c}^{(N)})\} \]

The learning task for a multi-label (multi-dimensional) classification paradigm

\[ \Omega_{\mathbf{x}} \equiv \Omega_{\mathbf{x}_1} \times \cdots \times \Omega_{\mathbf{x}_n} \xrightarrow{\phi} \Omega_{\mathbf{c}} \equiv \Omega_{\mathbf{c}_1} \times \cdots \times \Omega_{\mathbf{c}_d} \]

\[ \mathbf{x} \equiv (x_1, \ldots, x_n) \quad \rightarrow \quad \mathbf{c} \equiv (c_1, \ldots, c_d) \]

Overview of learning methods for multi-label classification

**Problem transformation methods**
- They transform the learning task into one or more single-label classification tasks
- They are algorithm independent
- (a) Transform to binary classification: binary relevance (Godbole and Sarawagi, 2004), classifiers chain (Read et al., 2011)
  - (b) Transform to multiclass classification: label powerset (Boutell et al., 2004), RAKEL (Tsoumakas et al., 2010)
  - (c) Identifying label dependencies: correlation based pruning (Tsoumakas et al., 2009), LBPR algorithm (Tenenboim et al., 2010)

**Algorithm adaptation methods**
- They extend specific learning algorithms in order to handle multi-label data directly
- Examples: classification trees (Clare and King, 2001), neural networks (Zhang and Zhou, 2006), k-nearest neighbors (Zhang and Hou, 2007), support vector machines (Elisseeff and Weston, 2001), random forest (Madjarov et al., 2012), Bayesian networks (van der Gaag and de Waal, 2006).
Mean accuracy and exact match for multi-label classification

<table>
<thead>
<tr>
<th>( \mathbf{x}^i )</th>
<th>( C_1 )</th>
<th>( C_2 )</th>
<th>( C_3 )</th>
<th>( C_4 )</th>
<th>( C_5 )</th>
<th>( \hat{C}_1 )</th>
<th>( \hat{C}_2 )</th>
<th>( \hat{C}_3 )</th>
<th>( \hat{C}_4 )</th>
<th>( \hat{C}_5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mathbf{x}^1 )</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>( \mathbf{x}^2 )</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>( \mathbf{x}^3 )</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>( \mathbf{x}^4 )</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( \mathbf{x}^5 )</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

- **Mean accuracy** computes the mean of the accuracies in each of the \( d \) class variables

\[
\text{Mean accuracy}(\phi) = \frac{1}{d} \sum_{j=1}^{d} \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}(\hat{c}_j^i = c_j^i)
\]

where \( \mathbb{I}(\text{true}) = 1 \) and \( \mathbb{I}(\text{false}) = 0 \)

In the table: Mean accuracy(\( \phi \)) = \( \frac{1}{5} (1 + 1 + 0.6 + 0.6 + 1) = 0.804 \)

- **Exact match** computes the fraction of correctly classified instances

\[
\text{Exact match}(\phi) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}(\hat{c}^i = c^i)
\]

In the table: Exact match(\( \phi \)) = \( \frac{1}{5} (0 + 1 + 1 + 0 + 0) = 0.4 \)
1. Multi-label classification and multi-dimensional classification

2. Binary relevance and label power set

3. Bayesian networks

4. Multi-dimensional Bayesian network classifiers

5. Applications

6. Conclusions

7. References
## Binary relevance (Godbole and Sarawagi, 2004)

Learns one binary classifier for each label independently of the rest of labels

Outputs the concatenation of their predictions

Does not consider label relationships

Any supervised classification method can be used, i.e., Bayesian network based classifiers

<table>
<thead>
<tr>
<th>( X )</th>
<th>( C_1 )</th>
<th>( C_2 )</th>
<th>( C_3 )</th>
<th>( C_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x^{(1)} )</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>( x^{(2)} )</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( x^{(3)} )</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>( x^{(4)} )</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( x^{(5)} )</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( X )</th>
<th>( C_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x^{(1)} )</td>
<td>1</td>
</tr>
<tr>
<td>( x^{(2)} )</td>
<td>0</td>
</tr>
<tr>
<td>( x^{(3)} )</td>
<td>1</td>
</tr>
<tr>
<td>( x^{(4)} )</td>
<td>0</td>
</tr>
<tr>
<td>( x^{(5)} )</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( X )</th>
<th>( C_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x^{(1)} )</td>
<td>0</td>
</tr>
<tr>
<td>( x^{(2)} )</td>
<td>0</td>
</tr>
<tr>
<td>( x^{(3)} )</td>
<td>0</td>
</tr>
<tr>
<td>( x^{(4)} )</td>
<td>1</td>
</tr>
<tr>
<td>( x^{(5)} )</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( X )</th>
<th>( C_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x^{(1)} )</td>
<td>0</td>
</tr>
<tr>
<td>( x^{(2)} )</td>
<td>1</td>
</tr>
<tr>
<td>( x^{(3)} )</td>
<td>0</td>
</tr>
<tr>
<td>( x^{(4)} )</td>
<td>0</td>
</tr>
<tr>
<td>( x^{(5)} )</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( X )</th>
<th>( C_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x^{(1)} )</td>
<td>1</td>
</tr>
<tr>
<td>( x^{(2)} )</td>
<td>1</td>
</tr>
<tr>
<td>( x^{(3)} )</td>
<td>1</td>
</tr>
<tr>
<td>( x^{(4)} )</td>
<td>0</td>
</tr>
<tr>
<td>( x^{(5)} )</td>
<td>0</td>
</tr>
</tbody>
</table>
Binary relevance with Bayesian network classifiers

\[ c^* = \arg \max_c p(c|x) = \arg \max_c p(x, c) \]

Taxonomy of discrete Bayesian network classifiers according to the factorization of \( p(x, c) \)

(Bielza and Larrañaga (2014))

- Naive Bayes
- Selective naive Bayes
- Semi-naive Bayes
- ODE
  - TAN
  - SPODE
- \( k \)-DB
- BAN
- Markov blanket-based
- Unrestricted
- Bayesian multinet

Decision boundary for discrete Bayesian network classifiers (Varando et al., 2015)
Binary relevance with Bayesian network classifiers

Naive Bayes (Minsky, 1961)

$$p(c|x) \propto p(c) \prod_{i=1}^{n} p(x_i|c)$$

A naive Bayes structure from which $p(c|x) \propto p(c)p(x_1|c)p(x_2|c)p(x_3|c)p(x_4|c)p(x_5|c)$
Binary relevance with Bayesian network classifiers

Tree-augmented naive Bayes (TAN) (Friedman et al. 1997)

Tree-augmented naive Bayes (TAN)

\[
p(c | x) \propto p(c)p(x_r | c) \prod_{i=1, i \neq r}^n p(x_i | c, x_j(i))
\]

where \( X_r \) denotes the root node and \( \{ X_{j(i)} \} = Pa(X_i) \setminus C \), for any \( i \neq r \)

(a) A TAN structure, whose root node is \( X_3 \), from which
\[
p(c | x) \propto p(c)p(x_1 | c, x_2)p(x_2 | c, x_3)p(x_3 | c)p(x_4 | c, x_3)p(x_5 | c, x_4);
\]
(b) Selective TAN (Blanco et al. 2005), from which
\[
p(c | x) \propto p(c)p(x_2 | c, x_3)p(x_3 | c)p(x_4 | c, x_3)
\]
Binary relevance with Bayesian network classifiers

**Markov blanket-based Bayesian classifier (Koller and Sahami, 1996)**

If $C$ has parents:

$$p(c|x) \propto p(c|pa(c)) \prod_{i=1}^{n} p(x_i|pa(x_i))$$

- The **Markov blanket of $C$** is the only knowledge needed to predict its behavior
- Bayesian classifiers based on identifying the Markov blanket of the class variable

![Diagram showing a Bayesian network with nodes $X_1, X_2, X_3, X_4, X_5$, where $C$ is the class variable, and its Markov blanket is $\{X_1, X_2, X_3, X_4\}$]

A **Markov blanket** structure for $C$, $\text{MB}_C = \{X_1, X_2, X_3, X_4\}$, from which

$$p(c|x) \propto p(c|x_2)p(x_1|c)p(x_2)p(x_3)p(x_4|c, x_3)$$
Each different set of labels becomes a different class in a new single-label classification task.

Most implementations of label powerset classifiers essentially ignore label combination that are not presented in the training set (cannot predict unseen labelsets).

Limited training examples for many classes.
Label powerset. Multiple diagnosis problem with probabilistic models. Peng and Reggia (1987a; 1987b)

\[
\begin{array}{c|cccc|cccc}
 & X_1 & \ldots & X_m & C_1 & \ldots & C_d \\
\hline
(x^{(1)}, c^{(1)}) & x_1^{(1)} & \ldots & x_m^{(1)} & c_1^{(1)} & \ldots & c_d^{(1)} \\
(x^{(2)}, c^{(2)}) & x_1^{(2)} & \ldots & x_m^{(2)} & c_1^{(2)} & \ldots & c_d^{(2)} \\
\vdots & \vdots & & \vdots & \vdots & & \vdots \\
(x^{(N)}, c^{(N)}) & x_1^{(N)} & \ldots & x_m^{(N)} & c_1^{(N)} & \ldots & c_d^{(N)} \\
\end{array}
\]

Optimal diagnosis as abductive inference: searching for the most probable explanation (MPE)

\[
(c_1^*, \ldots, c_d^*) = \arg \max_{(c_1, \ldots, c_d)} p(C_1 = c_1, \ldots, C_d = c_d | X_1 = x_1, \ldots, X_m = x_m)
\]

\[
= \arg \max_{(c_1, \ldots, c_d)} p(C_1 = c_1, \ldots, C_d = c_d) p(X_1 = x_1, \ldots, X_m = x_m | C_1 = c_1, \ldots, C_d)
\]

Number of parameters to be estimated: $2^d - 1 + 2^d(2^m - 1)$
Outline

1. Multi-label classification and multi-dimensional classification
2. Binary relevance and label power set
3. Bayesian networks
4. Multi-dimensional Bayesian network classifiers
5. Applications
6. Conclusions
7. References
### Bayesian networks (Pearl (1988) and Koller and Friedman (2009))

- **Years** $Y$ is the factory’s age, where $y$ denotes ‘more than 10 years’ and $\neg y$ denotes ‘less than 10 years’
- **Employees** $E$ represents the number of employees: more ($e$) or less ($\neg e$) than 100 employees
- **Machines** $M$ has two values: $m$ and $\neg m$ for ‘more than 20 machines’ and ‘less than 20 machines’
- **Pieces** $P$ also has two options: more ($p$) or less ($\neg p$) than 10,000 produced pieces per year
- **Failures** $F$ includes $f$ that stands for ‘more than two failures on average per month’ and otherwise the state is $\neg f$

<table>
<thead>
<tr>
<th>$Y$</th>
<th>$E$</th>
<th>$P$</th>
<th>$M$</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y$</td>
<td>$e$</td>
<td>0.70</td>
<td>0.10</td>
<td>0.85</td>
</tr>
<tr>
<td>$y$</td>
<td>$\neg e$</td>
<td>0.30</td>
<td>0.90</td>
<td>0.15</td>
</tr>
<tr>
<td>$\neg y$</td>
<td>$e$</td>
<td>0.20</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>$\neg y$</td>
<td>$\neg e$</td>
<td>0.80</td>
<td>0.95</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Hypothetical Bayesian network modeling factory production. To fully specify the JPD: $2^5 - 1 = 31$ parameters. The Bayesian network representing requires 17 input conditional probabilities.

The Bayesian networks factorizes the JPD as: $p(Y, E, M, P, F) = p(Y)p(E|Y)p(M|Y)p(P|E, M)p(F|M)$
Bayesian networks. Types of inference

Probabilistic reasoning: $p(X_i|e)$ where $E = e$ is the observed evidence

- **Abductive inference** finds the values of a set of variables that best explain the observed evidence
  - **Total abduction:** $\text{arg max}_U p(U|e)$, i.e., we find the most probable explanation (MPE),
  - **Partial abduction** solves the same problem for a subset of variables in $u$ (the explanation set), referred to as the partial maximum a posteriori (MAP)

- **Predictive reasoning:** we predict the effect (produced pieces) given a cause (machines), $p(p|m) = 0.62$ (in the figure) and $p(p|m)$ = 0.30 (not shown)

Inference on the factory production example. (a) Prior distributions $p(X_i)$. (b) $p(X_i|m)$

- **Diagnostic reasoning:** we diagnose the causes given the effects. Given the effect $F = f$, the probability of the cause being many machines is $p(m|f) = 0.86$

- **Intercausal reasoning:** if we know that the factory has many employees ($E = e$), this would explain the observed high number of pieces produced $p$ and would lower the probability of $\text{Machines} = m$ being the cause:

  
  $p(m|p, e) = 0.32 < p(m|p) = 0.45$
Bayesian networks. Learning parameters

(a) A Bayesian network structure with $|\Omega_{X_i}| = 2$, $i = 1, 3, 4$, $|\Omega_{X_2}| = 3$ and $q_1 = q_2 = 0$, $q_3 = 6$, $q_4 = 2$. (b) A dataset with $N = 6$ for $\{X_1, \ldots, X_4\}$ from which the structure in (a) has been learned

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1 = (\theta_{1-1}, \theta_{1-2})$</td>
<td>$(p(X_1 = 1), p(X_1 = 2))$</td>
</tr>
<tr>
<td>$\theta_2 = (\theta_{2-1}, \theta_{2-2}, \theta_{2-3})$</td>
<td>$(p(X_2 = 1), p(X_2 = 2), p(X_2 = 3))$</td>
</tr>
<tr>
<td>$\theta_3 = (\theta_{311}, \theta_{312}, \ldots, \theta_{361}, \theta_{362})$</td>
<td>$(p(X_3 = 1</td>
</tr>
<tr>
<td>$\theta_4 = (\theta_{411}, \theta_{412}, \theta_{421}, \theta_{422})$</td>
<td>$(p(X_4 = 1</td>
</tr>
</tbody>
</table>

- To estimate $\theta_{1-1} = p(X_1 = 1)$, we find four out of six instances in the $X_1$ column, and hence, $\hat{\theta}_{1-1}^{ML} = 2/3$
- To estimate $\theta_{322} = p(X_3 = 2|X_1 = 1, X_2 = 2)$, we find that neither of the two instances with $X_1 = 1, X_2 = 2$, include either $X_3 = 2$ or $\hat{\theta}_{322}^{ML} = 0$ (this is a case where $N_{ijk} = 0$)
- To estimate $\theta_{361} = p(X_3 = 1|X_1 = 2, X_2 = 3)$, we find that $\hat{\theta}_{361}^{ML}$ is undefined since there are no instances with $X_1 = 2, X_2 = 3$ (i.e., $N_{ij} = 0$). However, the Laplace estimates yield $\hat{\theta}_{322}^{Lap} = 1/4, \hat{\theta}_{361}^{Lap} = 1/2$
**Bayesian networks. Learning structures. Constraint-based methods**

- **Constraint-based methods** use the data to statistically test conditional independences among triplets of variables.
- The goal is to build a **DAG that represents** a large percentage (and whenever possible all) of the identified conditional independence constraints.
- The **PC algorithm** (Spirtes and Glymour, 1991) starts with all nodes connected by edges and follows three steps:
  1. Step 1 outputs the adjacencies in the graph, i.e., the **skeleton** of the learned structure.
  2. Step 2 identifies **colliders** (a collider or converging connection at node $X$ is $Y \rightarrow X \leftarrow Z$).
  3. Step 3 orients the edges and outputs the complete partially DAG (CPDAG), the Markov equivalence class of DAGs.
Bayesian networks. Score and search methods

Methods for Bayesian network structure learning based on score and search
Finding a network structure that optimizes a score is an **NP-hard** (Chickering, 1996)

Different search spaces:

- **Space of DAGs** whose cardinality was given by Robinson (1977):

\[
 f(n) = \sum_{i=1}^{n} (-1)^{i+1} \binom{n}{i} 2^{i(n-i)} f(n-i), \text{ for } n > 2
\]

with \( f(0) = f(1) = 1 \)

- **Space of Markov equivalence classes** is smaller than the space of DAGs. The \#DAGs/#CPDAGs ratio approaches an asymptote of about 3.7 (Gillispie and Perlman, 2002)

- **Space of orderings** of the variables: some learning algorithms (e.g., the K2 algorithm) only work with a fixed order. Cardinality: \( n! \)
Bayesian networks. Score and search methods. Scores

The need of penalized likelihood scores

- The estimated log-likelihood of the data given the BN:

\[
\log L(\hat{\theta}|D, G) = \log p(D|G, \hat{\theta}) = \log \prod_{i=1}^{n} q_i \prod_{j=1}^{R_i} \prod_{k=1}^{N_{ijk}} \hat{\theta}_{ijk}^{N_{ijk}} = \sum_{i=1}^{n} \sum_{j=1}^{q_i} \sum_{k=1}^{R_i} N_{ijk} \log \hat{\theta}_{ijk}
\]

where \(\hat{\theta}_{ijk} = \hat{\theta}_{ijk}^{ML} = \frac{N_{ijk}}{N_{ij}}\), the frequency counts in \(D\)

- This score increases monotonically with the complexity of the model
- The optimal structure would be the complete graph

Structural overfitting: the likelihood of the training data is higher for denser graphs, but it degrades for the test data
Penalized likelihood scores

- **General expression** of a family of penalized log-likelihood scores:

\[
Q^{\text{Pen}}(\mathcal{D}, \mathcal{G}) = \sum_{i=1}^{n} \sum_{j=1}^{q_i} \sum_{k=1}^{R_i} N_{ijk} \log \frac{N_{ijk}}{N_{ij}} - \text{dim}(\mathcal{G})\text{pen}(N)
\]

where \(\text{dim}(\mathcal{G}) = \sum_{i=1}^{n} (R_i - 1)q_i\) denotes the model dimension (number of parameters needed in the Bayesian network), and \(\text{pen}(N)\) is a non-negative penalization function.

- Depending on \(\text{pen}(N)\):
  - If \(\text{pen}(N) = 1\), the score is called Akaike's information criterion (Akaike, 1974)
  - If \(\text{pen}(N) = \frac{1}{2} \log N\), the score is the Bayesian information criterion (BIC) (Schwartz, 1978)
**Dynamic, temporal and continuous time Bayesian networks**

- **Dynamic Bayesian networks** (Dean and Kanazawa, 1989)

![Dynamic Bayesian network structure](image)

A dynamic Bayesian network structure with four variables $X_1$, $X_2$, $X_3$ and $X_4$ and three time slices ($T = 3$): (a) prior Bayesian network; (b) transition network, with the first-order Markovian transition assumption; (c) dynamic Bayesian network unfolded in time for three time slices

- **Temporal nodes Bayesian networks** (Galán et al., 2007)
- **Continuous time Bayesian networks** (Nodelman et al., 2002)
Bayesian networks. Software

<table>
<thead>
<tr>
<th></th>
<th>HUGIN(^1)</th>
<th>GeNi(^2)</th>
<th>Open-Markov(^3)</th>
<th>gRain(^4)</th>
<th>bnlearn(^5)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exact inference</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Junction tree</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Approximate inference</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probabilistic logic sampling</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td><strong>Constraint-based learning</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC algorithm</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Score+search</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>K2 algorithm</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
</tbody>
</table>

The symbol √ denotes that the inference or structure learning method is available in the respective software.

\(^1\) https://www.hugin.com/
\(^2\) https://www.bayesfusion.com/
\(^3\) http://www.openmarkov.org/
\(^4\) https://CRAN.R-project.org/package=gRain
\(^5\) http://www.bnlearn.com/
Outline

1. Multi-label classification and multi-dimensional classification
2. Binary relevance and label power set
3. Bayesian networks
4. Multi-dimensional Bayesian network classifiers
5. Applications
6. Conclusions
7. References
Multiple diagnosis problem. Direct approach

Optimal diagnosis as abductive inference: searching for the most probable explanation (MPE)

\[
(c_1^*, \ldots, c_d^*) = \arg \max_{(c_1, \ldots, c_d)} p(C_1 = c_1, \ldots, C_d = c_d | X_1 = x_1, \ldots, X_m = x_m)
\]

Number of parameters to be estimated for the case of binary predictors and classes:

\[2^d - 1 + 2^d(2^m - 1)\]
Multiple diagnosis with multi-dimensional Bayesian network classifiers (MBCs) (van der Gaag and de Waal, 2006)

- In a multi-dimensional Bayesian network classifier (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 $(d + m = n)$

- Three subgraphs in the structure of a multi-dimensional Bayesian network classifier
  - Class subgraph: $\mathcal{G}_C$
  - Bridge subgraph: $\mathcal{G}_{C \times X}$
  - Feature subgraph: $\mathcal{G}_X$
Examples of MBCs structures

(a) Empty-empty MBC

(b) Tree-tree MBC

(c) Polytree-DAG MBC
**Tractability of MPE in MBCs with class bridge decomposable MBCs (Bielza et al., 2011)**

MPE is generally **NP-hard** in Bayesian networks (Kwisthout, 2011)

An MBC is **class-bridge decomposable (CB-decomposable MBC)** (Bielza et al., 2011) if:

1. \( G_C \cup G_{CX} \) can be decomposed as: \( G_C \cup G_{CX} = \bigcup_{i=1}^{r} (G_{C_i} \cup G_{(CX)_i}) \) where \( G_{C_i} \cup G_{(CX)_i} \) with \( i = 1, ..., r \) are its maximal connected components

2. Non-shared children: \( Ch(\mathcal{V}_{C_i}) \cap Ch(\mathcal{V}_{C_j}) = \emptyset \), with \( i, j = 1, ..., r \) and \( i \neq j \), where \( Ch(\mathcal{V}_{C_i}) \) denotes the children of all the variables in \( \mathcal{V}_{C_i} \)

---

(a) A CB-decomposable MBC

(b) Its two maximal connected components
Theorem (Bielza et al., 2011)

Given a CB-decomposable MBC where $I_i = \prod_{C \in \mathcal{V}_C} \Omega_C$ denotes the sample space associated with $\mathcal{V}_C$, then

$$\max_{c_1, \ldots, c_d} p(C_1 = c_1, \ldots, C_d = c_d | X_1 = x_1, \ldots, X_m = x_m)$$

$$\propto \prod_{i=1}^r \max_{\mathcal{C} \subseteq \mathcal{V}_C} \prod_{c \in \mathcal{V}_C} p(c | \text{pa}(c)) \prod_{x \in \text{Ch}(\mathcal{V}_C)} p(x | \text{pa}_{\mathcal{V}_C}(x), \text{pa}_{\mathcal{V}_{\mathcal{X}}}(x))$$

where $c \downarrow \mathcal{V}_C$ represents the projection of vector $c$ to the coordinates found in $\mathcal{V}_C$.

Here $\mathcal{V}_C = \{ C_1, C_2, C_3 \}$, $\mathcal{V}_C = \{ C_4, C_5 \}$, $\text{Ch}(\mathcal{V}_C) = \{ X_1, X_2, X_3, X_4 \}$ and $\text{Ch}(\mathcal{V}_C) = \{ X_5, X_6 \}$. Note that $\text{Ch}(\mathcal{V}_C) \cap \text{Ch}(\mathcal{V}_C) = \emptyset$ as required in the theorem.
Learning MBCs from data. DAG-DAG MBC. Filter

**DAG–DAG MBC** (Bielza et al., 2011)

- **Score**: penalized likelihood or Bayesian score
- **Greedy search** in five steps:
  1. Learn the class subgraph
  2. Learn the feature subgraph
  3. Propose a candidate bridge subgraph
  4. Obtain a candidate feature subgraph
  5. Decide on bridge and feature subgraph candidates
**DAG–DAG MBC** (Bielza et al., 2011)

1. Learn the class subgraph
**DAG–DAG MBC** (Bielza et al., 2011)

2. Learn the feature subgraph and
3. Propose a candidate bridge subgraph
DAG–DAG MBC (Bielza et al., 2011)

2. Learn the feature subgraph and
3. Propose a candidate bridge subgraph
DAG–DAG MBC (Bielza et al., 2011)

4. Obtain a candidate feature subgraph and
5. Decide on bridge and feature subgraph candidates
Learning MBCs from data. DAG-DAG MBC. Wrapper

**DAG–DAG MBC** (Bielza et al., 2011)

\[ i = 0 \]

1. \( G^{(i)} = \emptyset \). \( Acc = Acc^{(i)} \)

2. Whenever there are arcs that can be added to \( G^{(i)} \) (and not previously discarded): Add one arc to \( G_{C}^{(i)} \), \( G_{C, \chi}^{(i)} \) or \( G_{\chi}^{(i)} \) and obtain the new \( G^{(i+1)} \) and \( Acc^{(i+1)} \)

3. If \( Acc^{(i+1)} > Acc^{(i)} \), \( Acc = Acc^{(i+1)} \), \( i = i + 1 \), and go to 2
   Else discard the arc and go to 2

4. Stop and return \( G^{(i)} \) and \( Acc \)
Learning CB decomposable MBCs from data (Borchani et al., 2010)

Phase I. Learn bridge subgraph

Starting from an empty graphical structure, learn a selective naive Bayes for each class variable.

Check the non-shared children property to induce an initial CB-decomposable MBC (remove all common children based on two criteria: feature insertion rank and accuracy).

The result of this phase is a simple CB-decomposable MBC where only the bridge subgraph is defined (class and feature subgraphs are empty).
Introduce the dependence relationships between the feature subgraph

Fix the maximum number of iterations

In each iteration an arc is selected at random between a pair of feature variables. If there is an accuracy improvement, the arc is added, otherwise it is discarded
Learning CB decomposable MBCs from data (Borchani et al., 2010)

Phase III. Merge maximal connected components

Merging the maximal connected components
- All possible arc additions between the class variables are evaluated, adding the arc improving the accuracy the most (from $r$ to $r - 1$ maximal connected components)
- A bridge update step is performed inside the new induced maximal connected component

Bridge subgraph update

Feature subgraph update
- Update the feature subgraph by inserting, one by one, additional arcs between feature variables
- This phase iterates over these three steps (stopping criteria: no more component merging can improve the accuracy or $r = 1$)
Markov blanket MBC (MB-MBC) learning algorithm

- Apply the HITON algorithm (Aliferis et al., 2003) to each class variable to determine Markov blankets for each of them.
- Given the MBC definition, direct parents of any class variable $C_i$, $i = 1, \ldots, d$, can only be among the remaining class variables, whereas direct children or spouses of $C_i$ can include either class or feature variables.
- The MBC subgraphs based on the results of the HITON algorithm:
  - **Class subgraph**: firstly insert an edge between each class variable $C_i$ and any class variable belonging to its corresponding parents-children set $PC(C_i)$. Then, direct all these edges using the PC algorithm’s edge orientation rules.
  - **Bridge subgraph**: this is built by inserting an arc from each class variable $C_i$ to every feature variable belonging to $PC(C_i)$.
  - **Feature subgraph**: for every feature $X$ in the set $MB(C_i) \setminus PC(C_i)$, i.e., for every spouse $X$, we insert an arc from $X$ to the corresponding common child given by $PC(X) \cap PC(C_i)$.
A multi-dimensional Bayesian network classifier tree (MBCTree) is a classification tree with MBCs in the leaves.

- An internal node of an MBCTree corresponds to a feature variable $X_i$, as in standard classification trees, and has a labelled branch to a child for each of its possible values.
- A leaf of an MBCTree is an MBC over all the class variables and those features not present in the path from the root to the leaf.

Wrapper approach (greedy search) guided by the exact match accuracy for learning MBCTrees.
Outline

1. Multi-label classification and multi-dimensional classification
2. Binary relevance and label power set
3. Bayesian networks
4. Multi-dimensional Bayesian network classifiers
5. Applications
6. Conclusions
7. References
Therapies for HIV-1 are combinations or cocktails of antiretroviral drugs.

We aim to gain insight into the different interactions between drugs and resistance mutations.

Reverse transcriptase inhibitors (RTIs) consist of two groups of antiretroviral drugs preventing HIV-1 replication: nucleoside and nucleotide reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs).

NRTIs consist of seven drugs: Abacavir (ABC), Didanosine (DDI), Emtricitabine (FTC), Lamivudine (3TC), Stavudine (D4T), Tenofovir (TDF), and Zidovudine (AZT).

NNRTIs considers three drugs: Efavirenz (EFV), Nevirapine (NVP), and Delavirdine (DLV).

A total of 38 mutations associated with resistance to RTIs: 22 associated with NNRTs and 16 with NNRTIs.
A cell with the HIV-1. Reverse transcription
Predicting human immunodeficiency virus (HIV) type 1 inhibitors (Borchani et al., 2013)

- We analyzed reverse transcriptase and protease data sets obtained from the online Stanford HIV-1 database (Rhee et al., 2003)
- Treatment histories from 2855 patients that received either NRTIs, NNRTIs or both
- The data set contained a total of 4884 samples
- The number of RTIs varies from 1 to 8 drugs: 17 samples with 1 RTI, 25 with 2 RTIs, 157 with 3 RTIs, 698 with 4 RTIs, 1852 with 5 RTIs, 1600 with 6 RTIs, 483 with 7 RTIs, and 56 with 8 RTIs

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean accuracy</th>
<th>Exact match</th>
</tr>
</thead>
<tbody>
<tr>
<td>MB-MBC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{max}_{CS} = 1$</td>
<td>0.7108 ± 0.0221</td>
<td>0.1151 ± 0.0466</td>
</tr>
<tr>
<td>$\text{max}_{CS} = 2$</td>
<td>0.7062 ± 0.0191</td>
<td>0.0881 ± 0.0403</td>
</tr>
<tr>
<td>$\text{max}_{CS} = 3$</td>
<td>0.7019 ± 0.0153</td>
<td>0.0780 ± 0.0363</td>
</tr>
<tr>
<td>$\text{max}_{CS} = 4$</td>
<td>0.6995 ± 0.0145</td>
<td>0.0701 ± 0.0336</td>
</tr>
<tr>
<td>$\text{max}_{CS} = 5$</td>
<td>0.6978 ± 0.0106</td>
<td>0.0646 ± 0.0241</td>
</tr>
<tr>
<td>Tree-Tree</td>
<td>0.6968 ± 0.0163</td>
<td>0.0364 ± 0.0101</td>
</tr>
<tr>
<td>DAG-DAG Filter</td>
<td>0.7074 ± 0.0063</td>
<td>0.0240 ± 0.0066</td>
</tr>
<tr>
<td>DAG-DAG Wrapper</td>
<td>0.7095 ± 0.0040</td>
<td>0.0291 ± 0.0008</td>
</tr>
<tr>
<td>CB-MBC</td>
<td>0.7261 ± 0.0113</td>
<td>0.0382 ± 0.0105</td>
</tr>
</tbody>
</table>

Estimated performance metrics (mean ± std. deviation) for RTIs data set
Predicting human immunodeficiency virus (HIV) type 1 inhibitors (Borchani et al., 2013)

The graphical structure of the MBC learnt using RTIs data set
EQ-5D health states from PDQ-39 in Parkinson (Borchani et al., 2012)

Parkinson disease motor symptoms
PDQ-39 and EQ-5D: quality of life instruments to measure the degree of disability in PD

39-item Parkinson’s Disease Questionnaire: a specific instrument

PDQ-39 captures patient’s perception of his illness covering 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

**PDQ-39 QUESTIONNAIRE**

*Please complete the following*

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

<table>
<thead>
<tr>
<th>Question</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. Had difficulty doing leisure activities which you would like to do?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Had difficulty looking after your home, e.g. DIY, household tasks?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Had difficulty carrying bags of shopping?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Had problems walking half a mile?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Had problems walking 100 yards?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Had problems getting around the house as easily as you would like?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Pedro Larrañaga

Multi-dimensional Bayesian network classifiers
European Quality of Life - 5 Dimensions: a generic instrument

EQ-5D is a generic measure of health for clinical and economic appraisal

Mobility
I have no problems in walking about
I have some problems in walking about
I am confined to bed

Self-care
I have no problems with self-care
I have some problems washing and dressing myself
I am unable to wash and dress myself

Usual activities (eg. work, study, housework, family or leisure activities)
I have no problems with performing my usual activities
I have some problems with performing my usual activities
I am unable to perform my usual activities

Pain/discomfort
I have no pain or discomfort
I have moderate pain or discomfort
I have extreme pain or discomfort

Anxiety/depression
I am not anxious or depressed
I am moderately anxious or depressed
I am extremely anxious or depressed
EQ-5D health states from PDQ-39 in Parkinson (Borchani et al., 2012)

Mapping PDQ-39 to EQ-5D

<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>

$\phi : (PDQ_1, ..., PDQ_{39}) \rightarrow (EQ_1, ..., EQ_5)$
488 Parkinson’s patients. Estimated measures over 5-fold cross-validation

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean accuracy</th>
<th>Exact match</th>
</tr>
</thead>
<tbody>
<tr>
<td>MB-MBC</td>
<td>0.7119 ± 0.0338</td>
<td>0.2030 ± 0.0718</td>
</tr>
<tr>
<td>CB-MBC</td>
<td>0.6807 ± 0.0285</td>
<td>0.1865 ± 0.0429</td>
</tr>
<tr>
<td>MNL</td>
<td>0.6926 ± 0.0430</td>
<td>0.1802 ± 0.0713</td>
</tr>
<tr>
<td>OLS</td>
<td>0.4201 ± 0.0252</td>
<td>0.0123 ± 0.0046</td>
</tr>
<tr>
<td>CLAD</td>
<td>0.4254 ± 0.0488</td>
<td>0.0143 ± 0.0171</td>
</tr>
</tbody>
</table>
EQ-5D health states from PDQ-39 in Parkinson (Borchani et al., 2012)

MB–MBC graphical structure

- Pain/discomfort
- Self-care
- Mobility
- Anxiety/depression

Nodes: pdq38, pdq12, pdq15, pdq10, pdq8, pdq3, pdq2, pdq1, pdq5, pdq6, pdq4, pdq7, pdq17, pdq18
Outline

1. Multi-label classification and multi-dimensional classification
2. Binary relevance and label power set
3. Bayesian networks
4. Multi-dimensional Bayesian network classifiers
5. Applications
6. Conclusions
7. References
MULTI-DIMENSIONAL BAYESIAN NETWORK CLASSIFIERS

- Specialized Bayesian networks for solving multi-label and multi-dimensional problems

**Advantages:**
- Transparency, interpretability
- Variety of inference (MPE) and learning algorithms
- Hierarchy of models organized by structural complexity
- Competitive results with the state-of-the-art methods when $d$ is about dozens

**Disadvantages:**
- Wrapper approaches for learning are very computationally demanding
- In large $d$ problems difficult to compete with state-of-the-art methods
Outline

1. Multi-label classification and multi-dimensional classification
2. Binary relevance and label power set
3. Bayesian networks
4. Multi-dimensional Bayesian network classifiers
5. Applications
6. Conclusions
7. References
References


References


Pedro Larrañaga

Multi-dimensional Bayesian network classifiers
References


BAYESIAN NETWORKS BASED
MULTI-DIMENSIONAL CLASSIFICATION

Pedro Larrañaga

Computational Intelligence Group
Artificial Intelligence Department
Technical University of Madrid

3rd International Conference on Big Data Analytics, Data Mining and Computational Intelligence
Madrid, July 18, 2018