Summary

- We compare human and mouse basal dendrites morphology by modelling them with Gaussian Bayesian networks.
- Such models can uncover differences in interactions among variables between the two species. They can be used to test hypotheses about the data.

Introduction

Comparing mouse and human neurons can give insight on translating our studies on rodents to humans. We can focus, for example, on the morphology basal dendrites of pyramidal cells.

It is common to compare features univariately across species. For example, human pyramidal cells have are larger than mouse pyramidal cells.

Bayesian networks

A Bayesian network (Koller and Friedman 2009; Bielza and Larrañaga 2014) can compactly encode a multivariate probability distribution. It leverages conditional independencies to factor the joint density into a product of local ones:

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

$$p(x_i | pa(x_i)) = \mathcal{N}(\theta_i + \beta_i x_{pa}(x_i), \sigma_i^2)$$

The d-separation criterion lets us test conditional independencies off the network. We can learn the network from data, with appropriate solutions for the general case. We can compute the probability, conditional to some event, of another event or set of variables, in a high-dimensional distribution, with exact and approximate inference algorithms.

Method

- If the data are scarce, we can assume they come from a multivariate Gaussian joint distribution. This implies that (1) the variables are marginally Gaussian; and (2) the dependencies among them are linear. While restrictive, this assumption is common, especially since scarce data prohibit complex models. We can test our assumption with Marild’s test (Marild 1970).
- We can learn, for example, the network structure $G$ with the tabu algorithm and the BIC score, which provides regularization. We can learn from 1000 bootstrap samples and use the arcs that appear in at least 80% of the networks (Scutari and Nagarajan 2011). We can perform approximate inference by sampling 100,000 instances from the learned networks for each query.

Results

- Correlated: $X_1$ and $X_5$, $X_1$ and $X_2$, $X_1$ and $X_6$, $X_1$ and $X_3$.
- Uncorrelated: $X_1$ and $X_6$ with the remaining variables.
- $X_1$ and $X_6$ marginally uncorrelated ($\rho = -0.13391$), uncorrelated in human given $X_6$, yet correlated in mouse given $X_6$.

Hypothetical example: empirical distribution of $X_1$ and $X_6$ (left), values of $X_1$ and $X_6$ sampled from human network when $X_1 > 0.3$ (center), and $X_1$ and $X_6$ sampled from the mouse network when $X_1 > 1.15$ and $X_5 = -0.2$ (right).

Testing hypotheses

The models indicate absence / presence of correlations, not their direction nor strength. For that, we need inference. When conditioning on above-average $X_6$ (right), $X_1$ decreases slightly in both mouse and human, but the relative difference is unchanged.

Discussion

- With more variables we could take better advantage of the Bayesian network formalism.
- Bayesian networks can account for clusters in data as well as hidden real-valued variables (factor analysis).

References


Acknowledgements

This project has received funding from the European Union’s Horizon 2020 Framework Programme for Research and Innovation under Specific Grant Agreement No. 785907 (HBP SGA2).