Multivariate comparison of human and mouse pyramidal cell dendritic morphologies

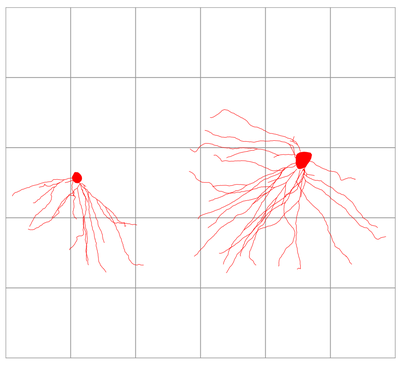
Bojan Mihaljević Concha Bielza Pedro Larrañaga
bmihaljevic@fi.upm.es
Computational Intelligence Group, Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid

Summary

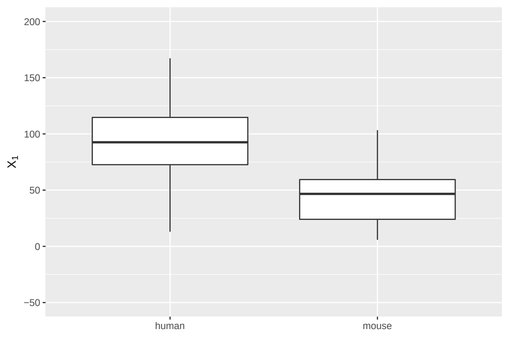
- We compare human and mouse basal dendrites morphology by modelling them with Gaussian Bayesian networks.
- Such models can uncover differences in interactions among variables between the two species. They can be used to test hypotheses about data.

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.



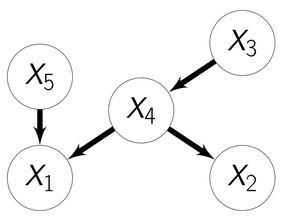
This ignores interactions among variables. Are such interactions similar across the two species? Is, for example, branch length difference more pronounced in cells with larger average bifurcation angles? A general approach to answering such questions is to model the joint probability distribution with a graphical model.

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:

$$p_{\mathcal{G}}(\mathbf{x}) = \prod_{i=1}^{n} p_{\mathcal{G}}(x_i \mid \mathbf{pa}_{\mathcal{G}}(x_i))$$

$$p_{\mathcal{G}}(x_i \mid \mathbf{pa}_{\mathcal{G}}(x_i)) = \mathcal{N}(\beta_i^0 + \boldsymbol{\beta}_i^T \mathbf{pa}_{\mathcal{G}}(x_i); \sigma_i^2)$$

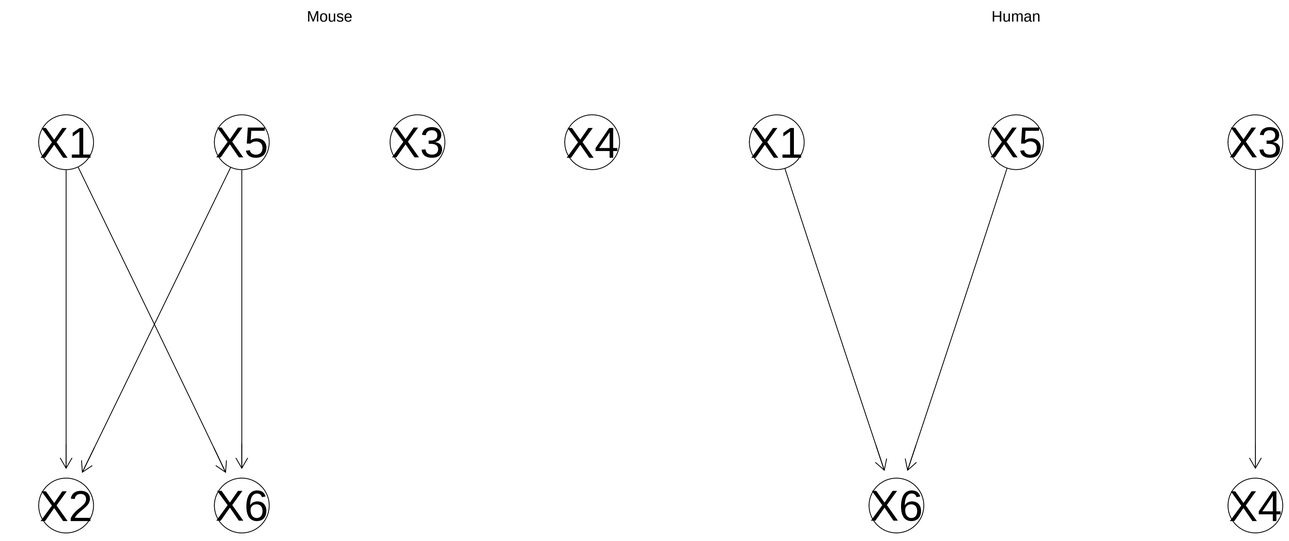


The d-separation criterion lets us read conditional independencies off the network. We can learn the network from data, with approximate solutions for the general case. We can compute the probability, conditional to some event, of another event or set of variables, in a high-dimensional distribution,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

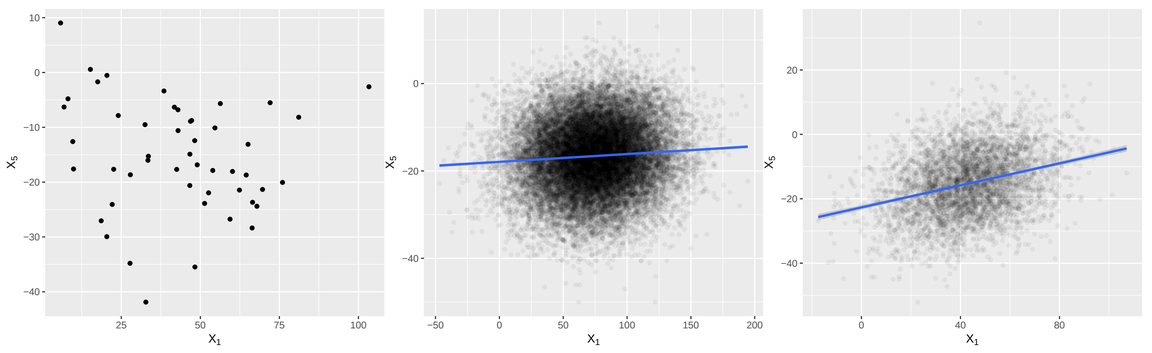
Method

- 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 prohibits complex models. We can test our assumption with Mardia's test (Mardia 1970).
- We can learn, for example, the network structure $\mathcal{G}$ 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 neworks for each query.

Results



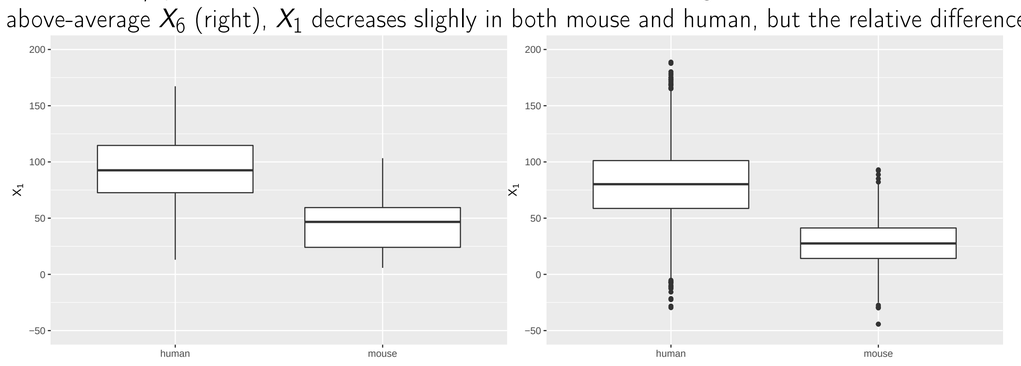
- Correlated: X_1 and X_6; X_5 and X_6; X_2 and X_6; X_2 and X_1;
- Uncorrelated: X_3 and X_4 with the remaining variables.
- X_1 and X_5 marginally uncorrelated (ρ = -0.153391); uncorrelated in human given X_6; yet correlated in mouse given $X_6$and X_2.



Hypothetical example: empirical distribution of X_1 and X_5 (left); values of X_1 and X_5 sampled from human network when $X_6 > 0.3$ (center); and X_1 and X_5 sampled from the mouse network when $X_2 > 1.15$‘ and $X_6 > -0.2$ (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 slighly 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

Bielza, Concha, and Pedro Larrañaga. 2014. "Bayesian Networks in Neuroscience: A Survey." *Frontiers in Computational Neuroscience* 8: 131.

Koller, Daphne, and Nir Friedman. 2009. *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA, USA: MIT press.

Mardia, Kanti V. 1970. "Measures of Multivariate Skewness and Kurtosis with Applications." *Biometrika* 57 (3). Oxford University Press: 519–30.

Scutari, Marco, and Radhakrishnan Nagarajan. 2011. "On Identifying Significant Edges in Graphical Models." In *Proceedings of Workshop on Probabilistic Problem Solving in Biomedicine. Bled, Slovenia: Springer-Verlag*, 15–27.

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).