

## Project: Multivariate analysis of Neuroimaging data using Latent variables

The last decades of neuroimaging research have brought immense insight into our understanding from the human brain function and structure. However, these findings suffer from small and unrepresentative samples. The [Human Connectome Project](http://www.humanconnectome.org/)<sup>1</sup> (HCP) and [UK Biobank](http://www.ukbiobank.ac.uk/)<sup>2</sup> have been launched recently to understand the biological basis of neurological and psychiatric illnesses. These projects collect multiple neuroimaging modalities, genomic data and numerous environmental factors, neurocognitive scores and clinical data on each subject, giving enormous opportunity to understand the function and structure of human brain. However, multivariate analysis of all data modalities remains largely rare.

Latent variable models where the high dimensional data is modelled with few unobserved low dimension factors can be used for joint analysis of all available data. GFA<sup>2</sup> and BASS<sup>3</sup> are the two software packages among many others that can fit Bayesian latent variable models.

The aim of this project is to walk through a recent publication<sup>1</sup> on multivariate analysis of UK Biobank data and find out the limitations of the study and suggest an alternative analysis strategy.

### Tasks:

Miller (2016) performed univariate and multivariate analyses on UK Biobank data to find association between imaging and behavioural using correlation and canonical correlation analysis (CCA). Perform the following tasks:

1. Read the paper by Miller and criticize the statistical analysis.
2. Understand the methodology behind GFA and BASS and
3. Perform Bayesian CCA on the dataset provided using BASS and GFA
4. Extend the analysis in (2) into multi-view settings where different imaging modality is considered as a view; compare the outputs and provide a thorough report
5. Think about how the outputs from the previous steps (2& 3) can be used for predictions or linear modelling.

### References:

1. Miller KL, Alfaro-Almagro F, Bangerter NK, et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci.* 2016;19(11):1523-1536.
2. Klami, A., Virtanen, S., & Kaski, S. (2013). Bayesian canonical correlation analysis. *Journal of Machine Learning Research*, 14(Apr), 965-1003.
3. Zhao, S., Gao, C., Mukherjee, S., & Engelhardt, B. E. (2016). Bayesian group factor analysis with structured sparsity. *The Journal of Machine Learning Research*, 17(1), 6868-6914.

---

<sup>1</sup> <http://www.humanconnectome.org/>

<sup>2</sup> <http://www.ukbiobank.ac.uk/>