# Machine learning for fMRI in Python: inverse inference with scikit-learn

The scikit-learn (Pedregosa 2010) is a general-purpose package for simple and efficient machine learning in Python. It can easily be combined with the various powerful Python packages for fMRI data processing, such as nipy (Brett 2007) or nibabel (Hanke 2009), to apply state-of-the-art learning methods to inverse inference, i.e. the prediction of behavior from brain imaging. The variety of supervised and unsupervised methods available in the scikit-learn as well as their performance greatly facilitates the use of machine learning for brain mapping. In addition, algorithms in the scikit-learn can be used as building-blocks to develop fMRI-specific statistical learning methods.

In the supervised learning settings, of particular importance for the brain-imaging community are the SVM-based methods and the penalized GLMs. For SVMs, the scikit-learn features libSVM bindings -amongst the richest and most performant- as well as associated methods such as recursive feature elimination. Penalized-GLM methods are available, to inject multivariate priors in a GLM, for regression or for classification (logistic regression); in particular L1 penalization and ARD priors are available to learn sparse predictive spatial maps. The scikit-learn also offers tools for cross validation and model selection to automatically set parameters or compare methods. For this, it can make use of multiple processors in a same computer. Unsupervised methods include clustering algorithms, mixture modeling as well as PCA and ICA.

Special care is taken across the scikit-learn code to present the various methods with a uniform interface, in order to make them easily interchangeable and to facilitate comparisons.

Finally, a documentation is available on the web site with a large number of examples and a special attention to ease of read for non machine-learning specialists. The scikit-learn heavily relies on the numpy package, which makes it possible to use the Python language as an array-based scientific computing environment, similar to matlab. Thus the scikit-learn can easily be picked up by matlab users who do not know Python.

We have implemented various fMRI-specific methods using the scikit-learn (Michel 2011a, Michel 2011b) and have used the scikit-learn to compare them to standard state-of-the-art learning methods. A full second-level inverse-inference pipeline, starting from first-level beta maps saved as Nifti file and extracting predictive brain features as well as cross-validated prediction scores can be written in 50 lines of code (Gramfort 2010). In addition, we use unsupervised learning methods to model resting-state functional connectivity, using ICA (Varoquaux 2010a) or covariance estimation (Varoquaux 2010b).

By bridging the gap between state-of-the-art machine-learning research and its applications to brain imaging, the scikit-learn can be used for multivariate brain mapping with high numerical and prediction performance, but also to develop new inverse-inference methods specially crafted to answer the needs and features of neuroimaging.

## References

- Brett, M. (2007), 'nipy: NeuroImaging in Python', http://nipy.sourceforge.net/
- Gramfort, A. (2010), 'NeuroImaging with the Scikit-learn: fMRI inverse inference tutorial', http://nisl.github.com/
- Hanke, M. (2009), 'nibabel: Access a cacophony of neuro-imaging file formats', http://nipy.sourceforge.net/nibabel/
- Michel, V. (2011a), 'Total variation regularization for fMRI-based prediction of behaviour', Trans. in Med. Im. in press
- Michel, V. (2011b), 'A supervised clustering approach for fMRI-based inference of brain states', Pattern recognition, in press
- Pedregosa, F. (2010), 'scikit-learn: Machine Learning in Python', http://scikit-learn.sourceforge.net/

- Varoquaux, G. (2010a), 'A group model for stable multi-subject ICA on fMRI datasets', Neuroimage 51, pp288-299
- Varoquaux, G. (2010b), 'Detection of brain functional-connectivity difference in post-stroke patients using group-level covariance modeling', MICCAI