Neuroimaging-specific machine learning

Brainhack 2019

Lisa Ronan

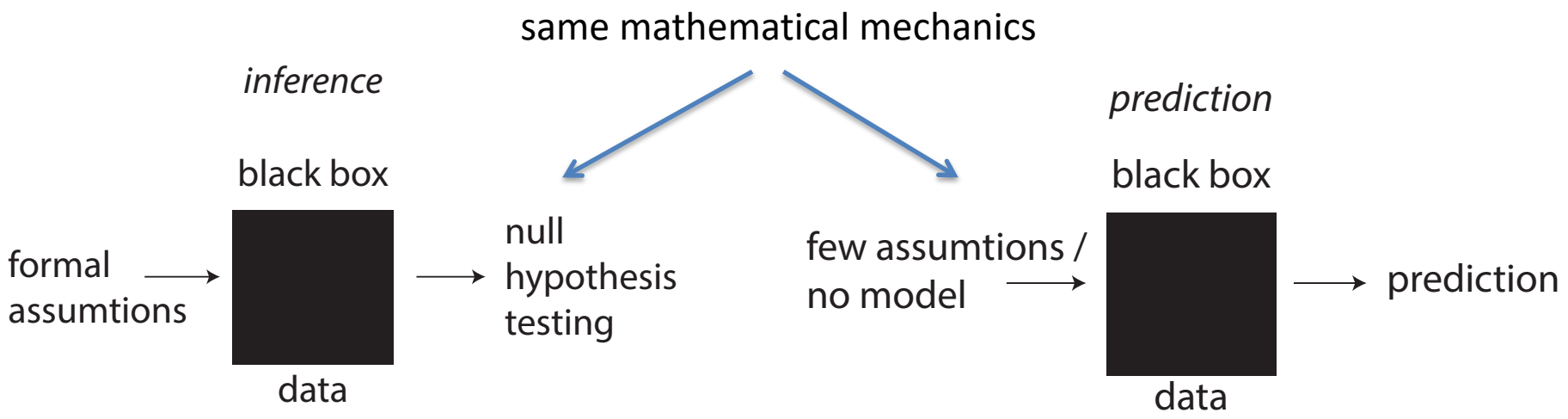
Prediction vs. inference

Model-driven classical hypothesis testing

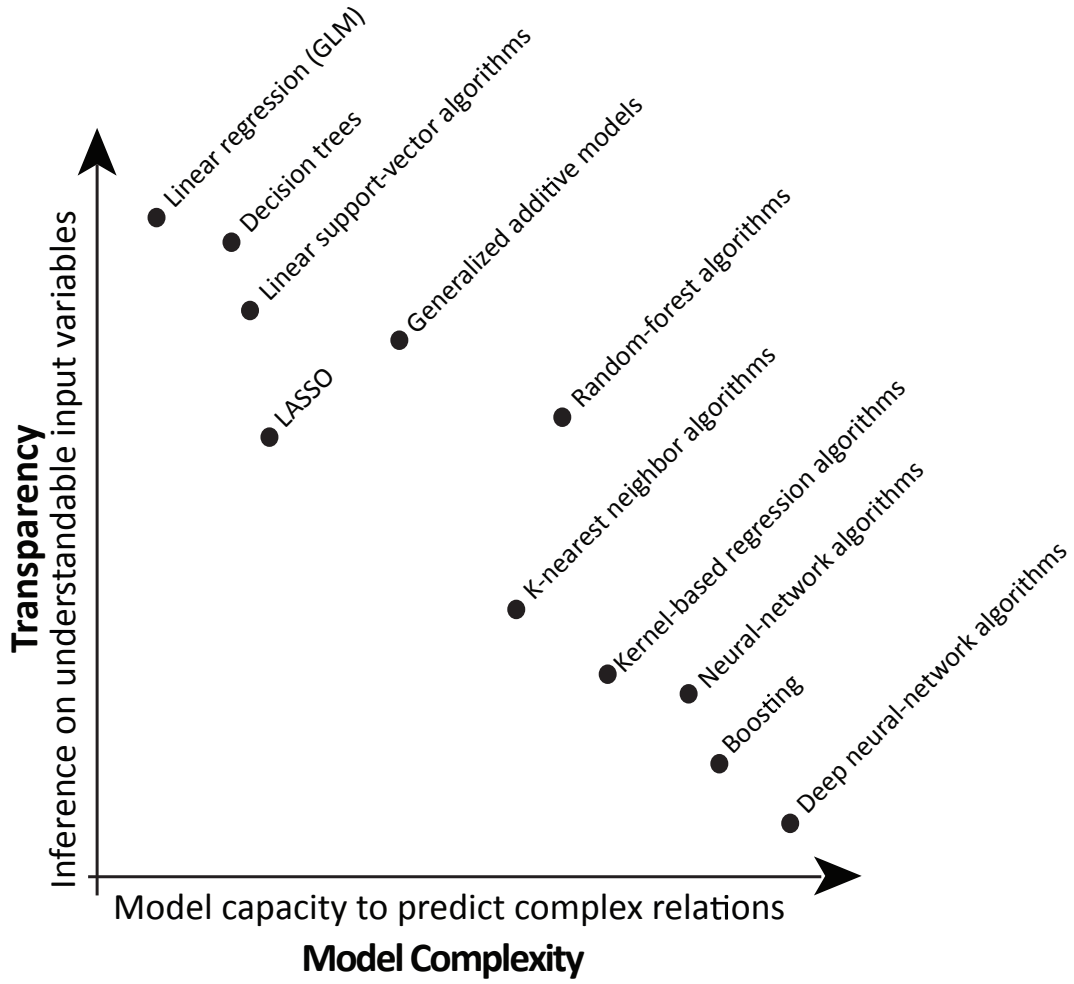
- Inference asks "*which input variable within a given dataset is an important contributor to the outcome? (or is a relatively more important contributor compared to other input variables?).*
- Typically uses whole data sample; not make single predictions
- Designed to aid mechanistic understanding

Data-driven learning algorithms

- Modeling for prediction typically asks a more heuristic type of question: *which data are collectively useful to distinguish individuals with or without the behavioral trait?"*
- Typically divides data in to "training" and "test" sets to evaluate model; can make predictions for single case.
- Not well suited to aid mechanistic understanding



Which algorithms?



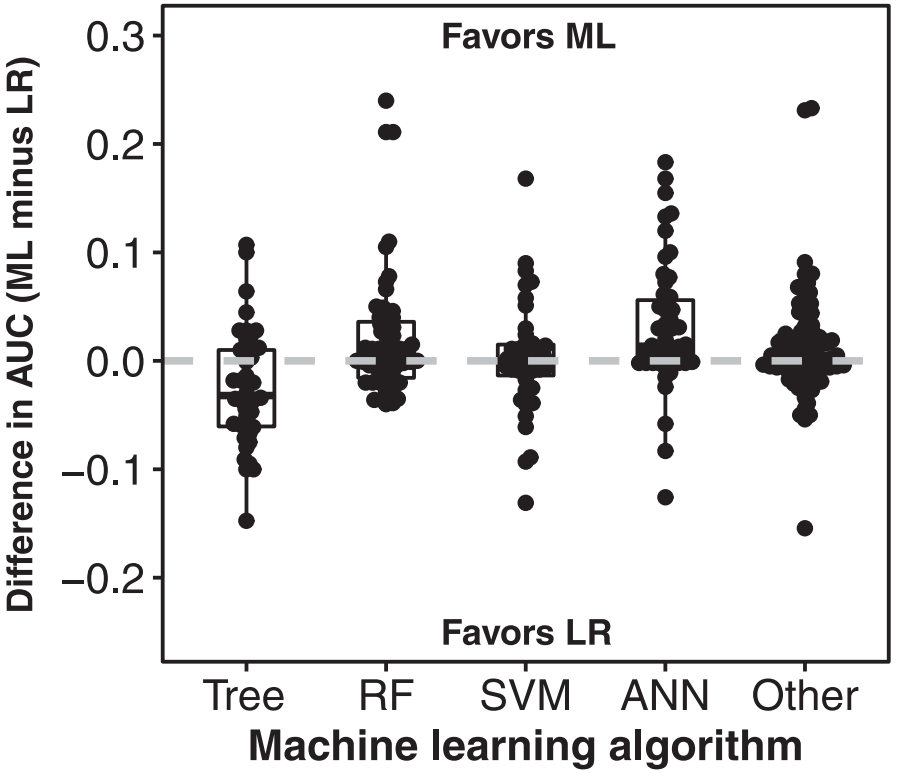
Bzdok & Ioannidis, TINS, 2019

"The superiority of modeling complex patterns over simple linear approaches should not be taken for granted, and merits case-by-case evaluation"
Bzdok and Ioannidis, 2019.

"Deep learning is only beneficial if the data have nonlinear relationships and if they are exploitable at currently available sample sizes" (Schulz et al., 2019).

Success typically on sample sizes of n > 1,000,000

LR vs. Deep Learning: Medical Records



Christodoulou et al., 2019

Meta analysis: 71 studies contrasing logistic regression vs. machine learning (including deep learning).

Median sample size 1,250 (72 – 4,000,000).

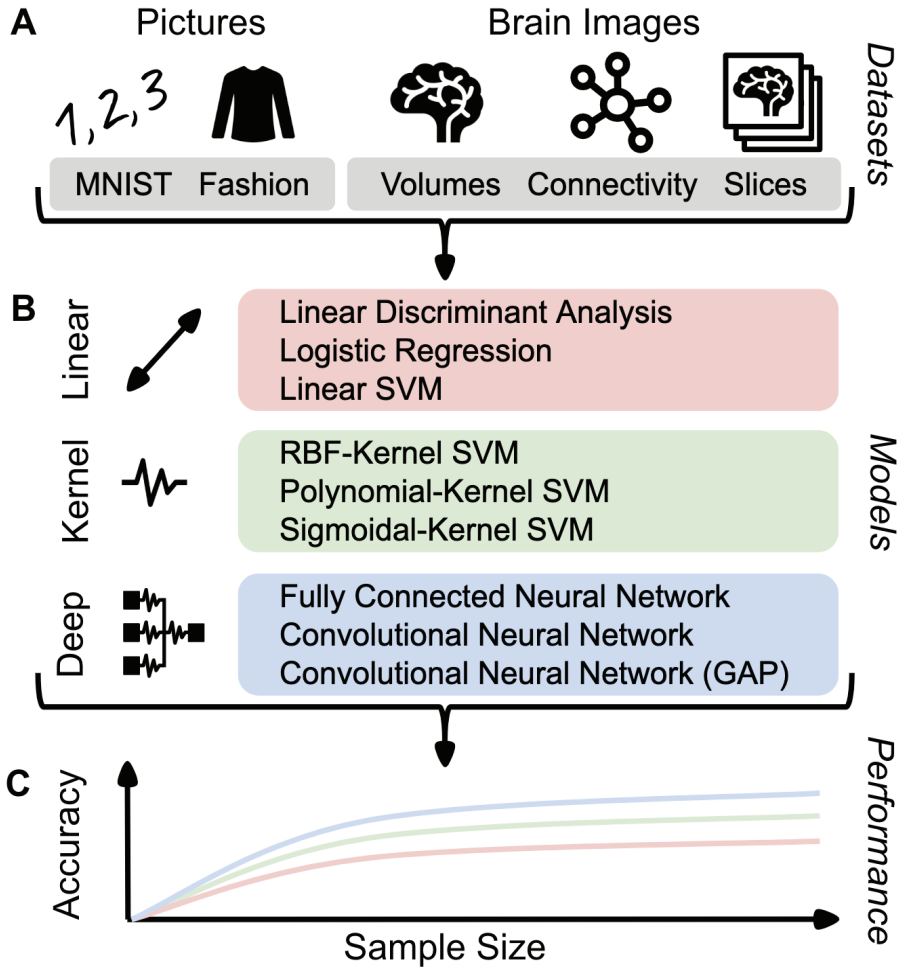
No. predictors: 5 – 563

Conclusion: *We found no evidence of superior performance of ML over LR.*

Deep learning for brains?

Deep learning for brains? Different linear and nonlinear scaling in UKBiobank brain images vs. machine-learning datasets

Schulz et al., 2019

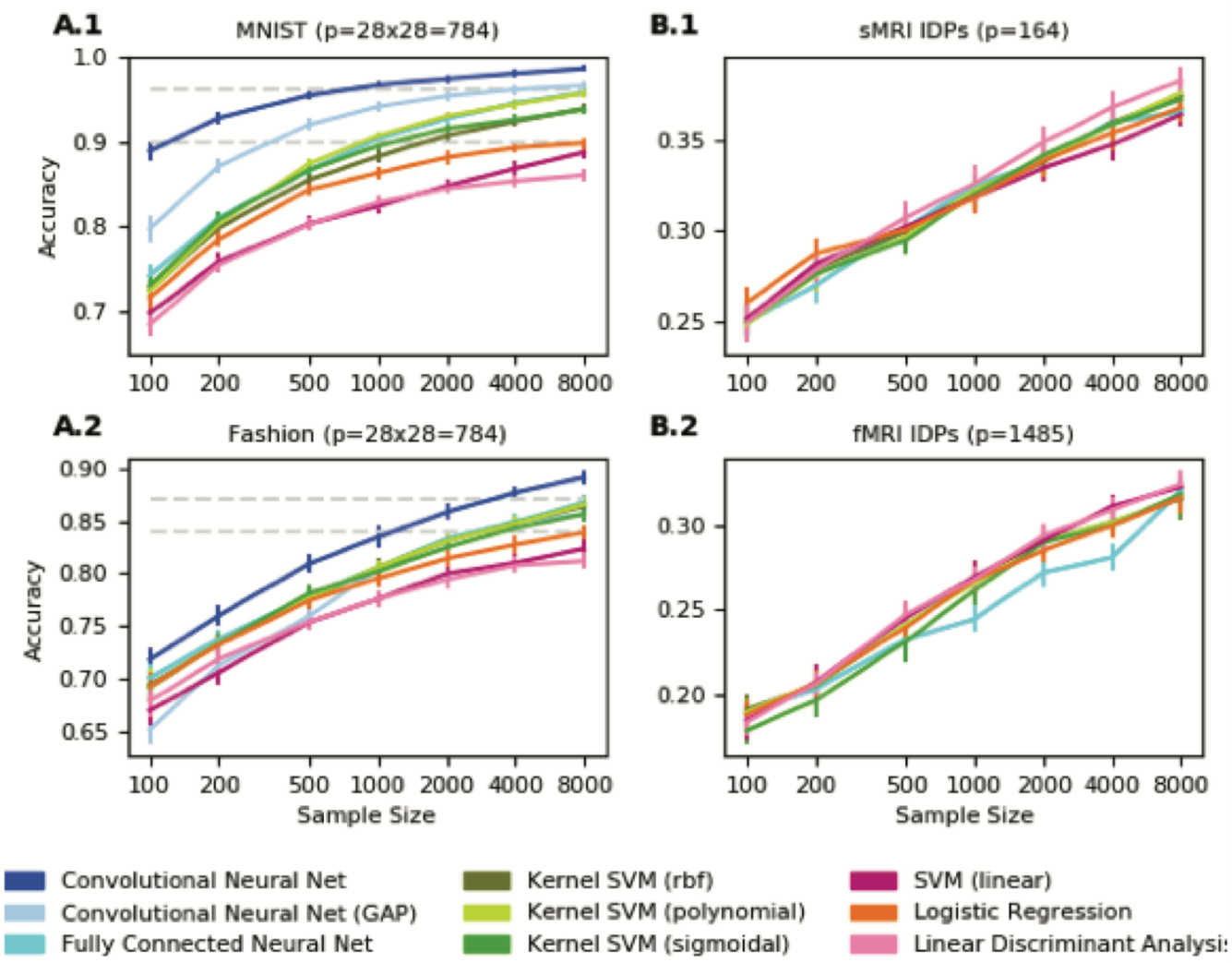


Experiment to test performance of linear vs. kernel vs. deep learning models on structural and functional UKBiobank MR brain data. Contrasted with the MINST and 'Fashion-MINST' datasets.

Contrasted Linear, kernel and deep-learning algorithms.

Evaluated performance on classification accuracy as well as change in accuracy as sample size increases.

Deep learning for brains?



Results: For MINST, there was a significant improvement in prediction accuracy escalating from linear (95%) to shallow non-linear (97%) to deep learning models (99%).

In contrast, linear models performed on par with more complex models in predicting age and sex.

Deep learning for brains?

Conclusion: "*Nonlinearities of common brain scans remain largely inaccessible to both kernel and deep learning methods at any examined scale" Schulz et al., 2019.*

In other words, we do not yet have the capacity to take advantage of more complex methods when modeling brain data.

Minimizing the Cost Function

The goal is to predict a target variable using a set of features, e.g. MR imaging data, and/or age, sex, etc.

In terms of algorithms, the goal is to find the optimal coefficients that minimise the *loss function*, i.e. the difference between the data and the model. There are different loss functions for different models, e.g.

$$MSE = \sum_{i=1}^{n} (y - \hat{y})^2$$

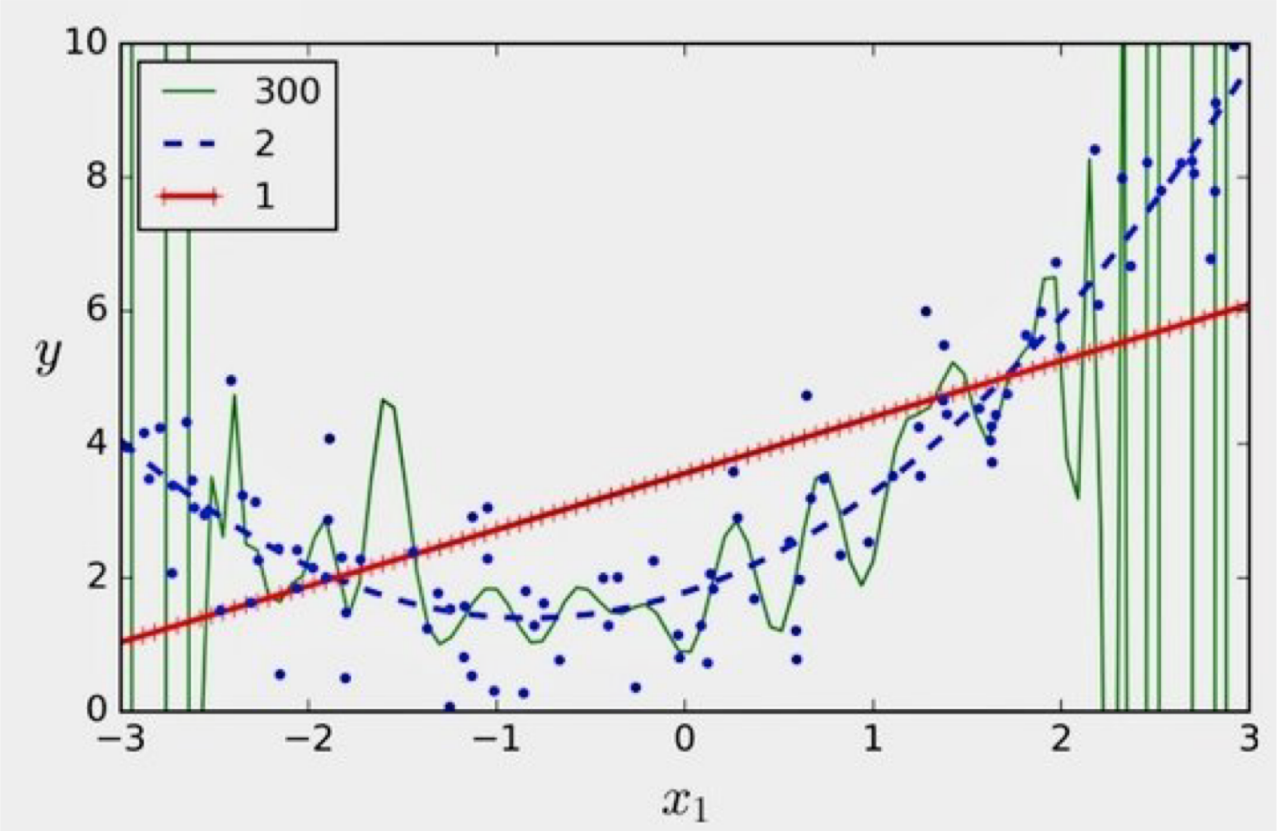
One weight is attributed per input feature.

However… while it is possible to accurately model the data, e.g. with a high-degree polynomial, these models will not generalize well to new data and thus will be poor predictors. This is called "*overfitting".*

The trade off between minimizing the cost function and avoiding overfitting is managed by regulariztion.

Regularization

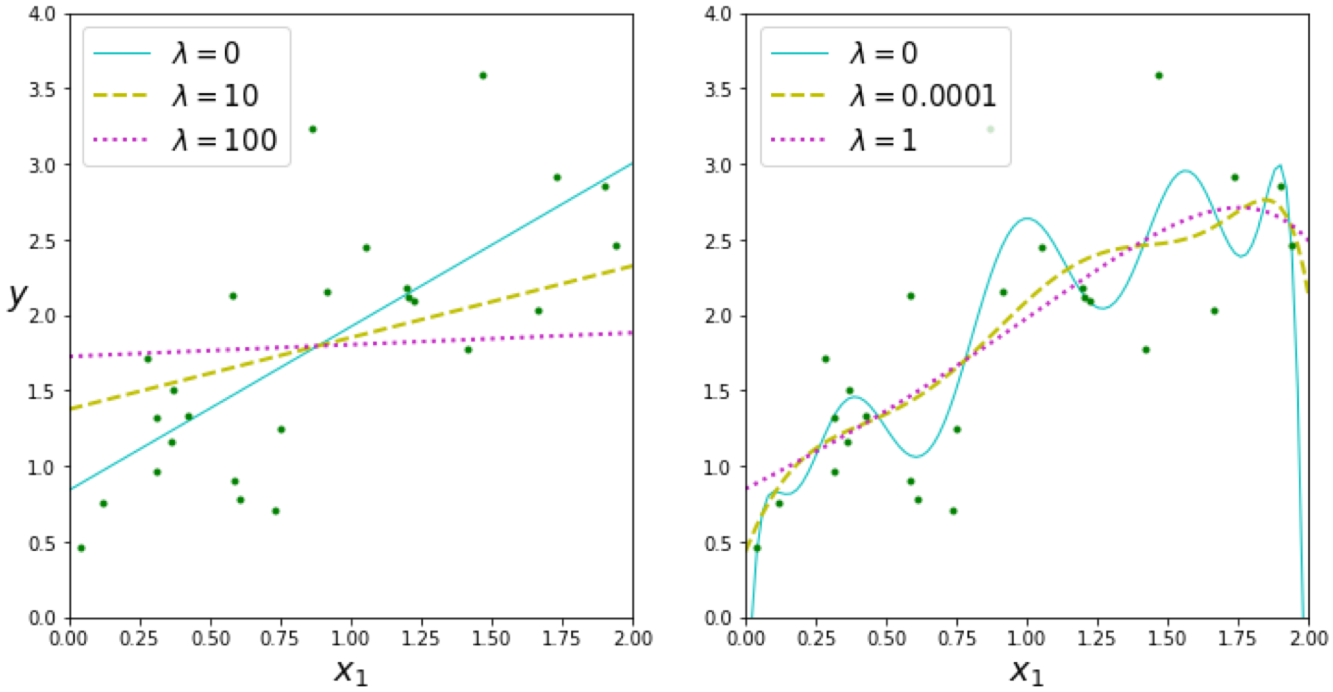
Regularization helps to minimise the problem of *overfitting*, that is, where both the relationship between variables and the noise in the data are both modeled.



Hands-on Machine Learning with Scikit-Learn and Tensor-Flow, A Géron

Regularization

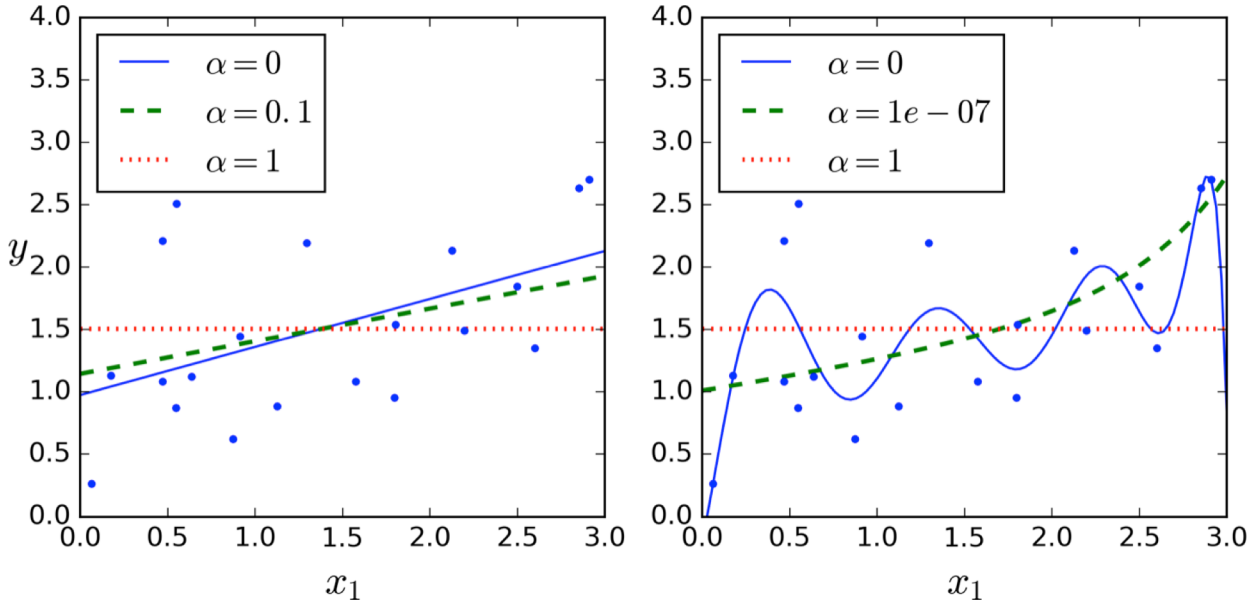
Ridge regression – imposes *L2* penalty so that solutions with large coefficients become unattractive. Essentially it minimises model weights.



Hands-on Machine Learning with Scikit-Learn and Tensor-Flow, A Géron

Regularization

Lasso regression – feature reduction - ensures that only a few features weights are non-zero. This not only reduces overfitting, but also increases interpretability.



Hands-on Machine Learning with Scikit-Learn and Tensor-Flow, A Géron

Ridge + Lasso = Elastic Net

Elastic Net encourages grouping, where few, strongly correlated predictors are included in the model.

Structured Sparsity

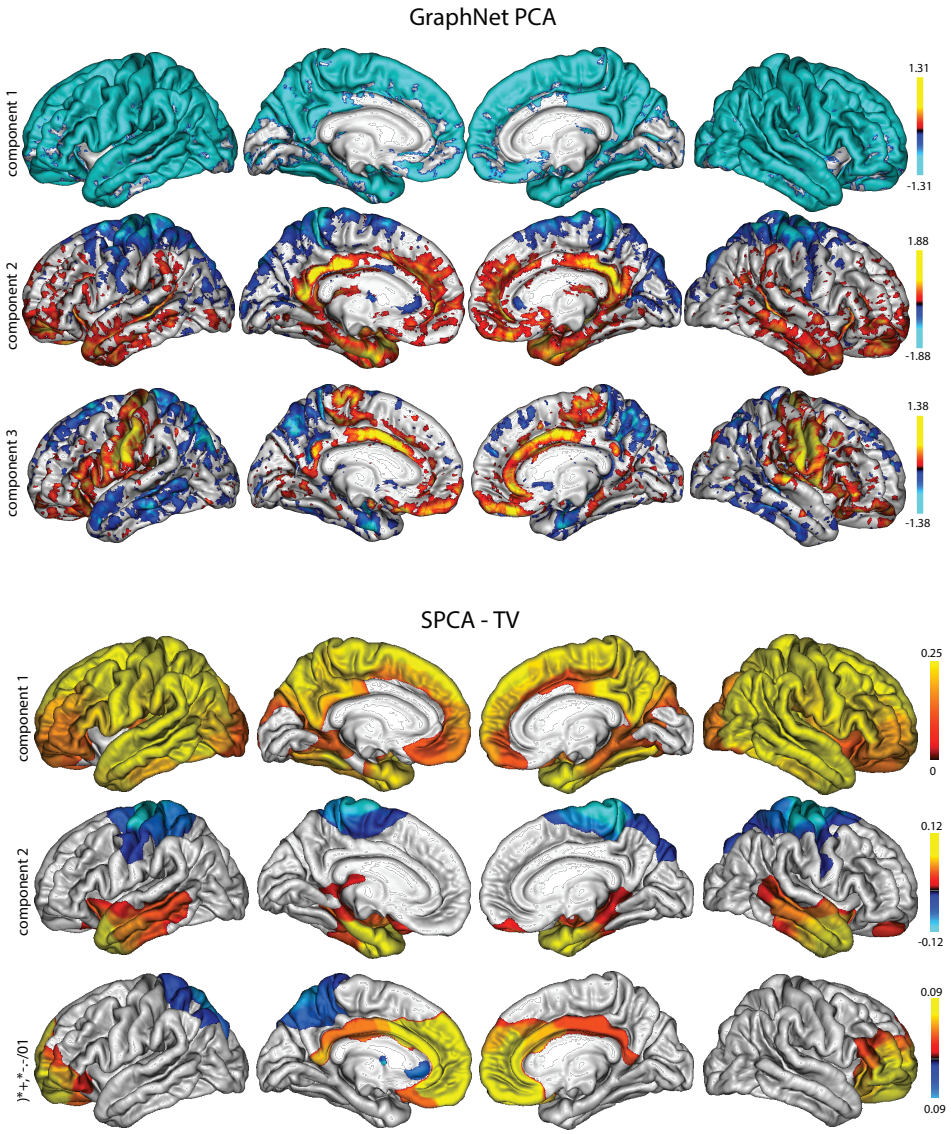
While Elastic Net is useful in generating sparse models, it does not take in to account the ***spatial structure of the data***, rather it treats all neuroimaging features as independent predictors – not realistic.

One solution is to implement "***structured sparsity***" which forces adjacent vertices to have similar weights by penalizing the pairwise differences between coefficients that are adjacent. This *GraphNet penalty* effectively generates smooth maps but does not ensure piece-wise smoothness.

An alternative approach is to use the *TV penalty (de Pierrefeu et al., 2018, 2019)* which does generate piecewise smoothness, effectively identifying predictive regions.

Structured sparsity not only makes use of the natural structure in the data, but also increases ***interpretability*** of the predictor maps.

GraphNet vs. TV penalty



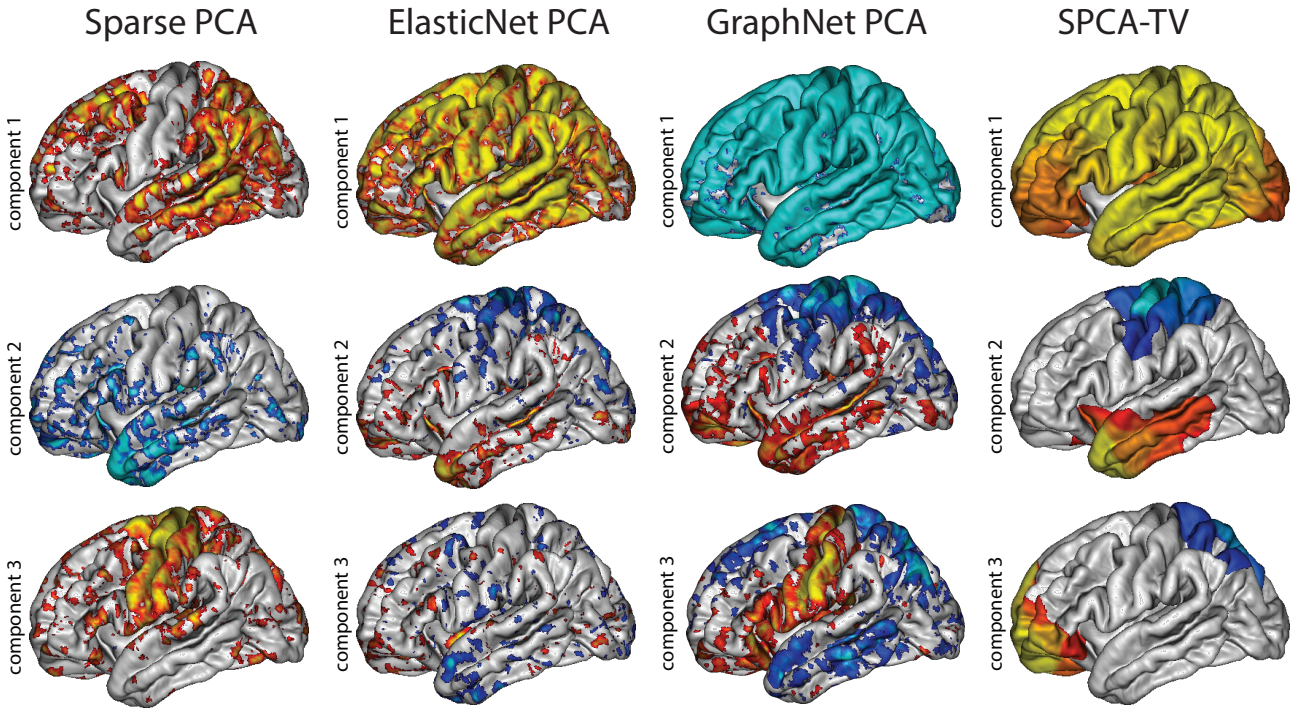
Loading vectors from 133 MCI patients
(de Pierrefeu et al., 2018)

GraphNet – smooth, not piecewise

TV – piecewise smooth

Machine Learning with Structured Sparsity: application to Neuroimaging-based Phenotyping in Autism Spectrum Disorder and Schizophrenia. A De Pierrefeu 2018

PCA in MCI

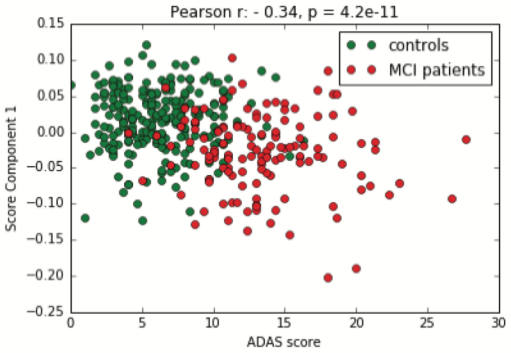


PCA methods applied to 133 patients with mild cognitive impairment (MCI) who converted to AD within two years during follow-up (ADNI dataset).

PCA used to identify patterns of atrophy (cortical thickness) explaining the variability in this population.

Method	Test Data Reconstruction Error	Dice Index
Sparse PCA	2991.8***	0.44**
ElasticNet PCA	2832.6**	0.43**
GraphNet PCA	2813.6*	0.62*
SPCA-TV	2795	0.65

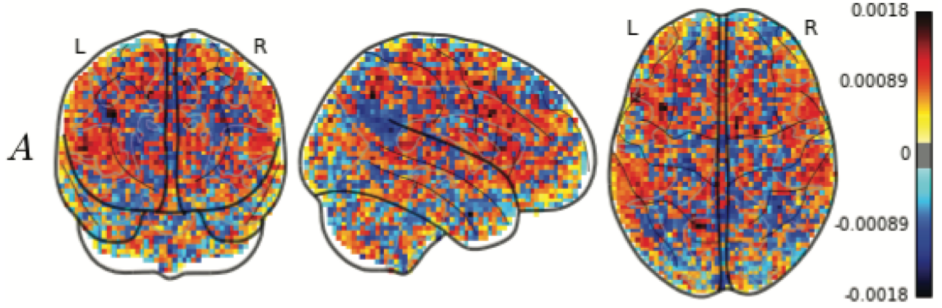
Scores averaged across 5 folds. Tested whether averaged scores obtained with existing PCA methods were significantly lower than scores from SPCA-TV.
*** $p < 10^{-3}$; ** $p < 10^{-2}$; * $p < 10^{-1}$



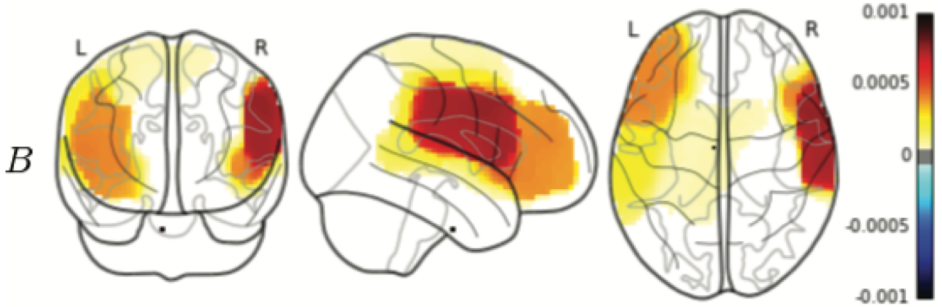
Machine Learning with Structured Sparsity: application to Neuroimaging-based Phenotyping in Autism Spectrum Disorder and Schizophrenia. A De Pierrefeu 2018

Prediction of activation patterns preceding hallucinations in schizophrenia

SVM (l2)



LR (Enet-TV)



Classifier	AUC	Accuracy	Specificity	Sensitivity
SVM	0.73*	0.73	0.78	0.67
Enet-TV	0.79*	0.74	0.76	0.71

Sensitivity (recall rate of "trans"; specificity recall rate of "off". Balanced accuracy = (Sen + Spe) / 2. AUC area under the curve.
*** $p < 10^{-3}$; ** $p < 10^{-2}$; * $p < 10^{-1}$

N = 37 subjects with schizophrenia, experienced average of 5.6 hallucinations per scan.

Dataset of 376 samples: 166 in resting-state ; 210 in prehallucination state, with 67,655 features (voxels).

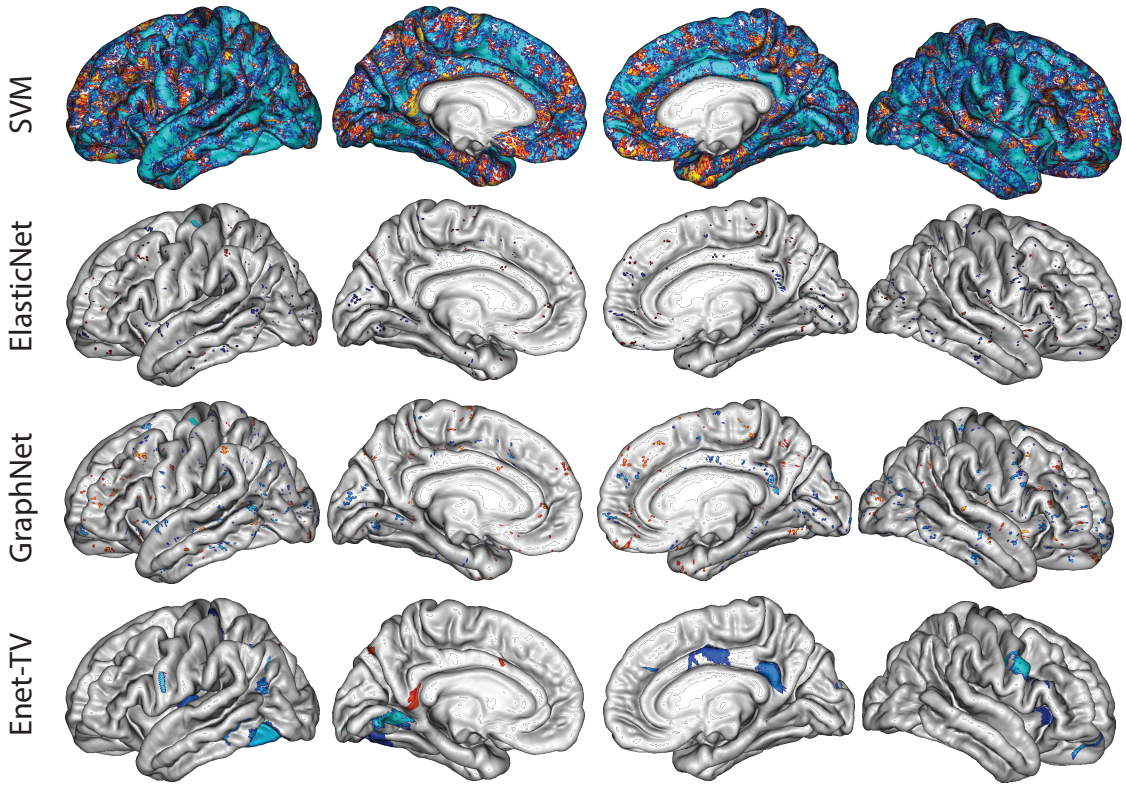
Classification of resting state vs. transition towards hallucinations.

Sensitivity: ability to identify transition to hallucinations

Specificity: ability to identify resting-state "off".

Prediction of activation patterns preceding hallucinations in patients with schizophrenia using machine learning with structured sparsity. De Pierrefeu (2018)

The neuroanatomical signature of schizophrenia



Classifier	AUC	Accuracy	Specificity	Sensitivity	r
SVM	0.69*	0.64*	0.63	0.65	
Enet-TV	0.7*	0.66*	0.6	0.71	0.76

R mean correlation between pairs of weight maps computed across 4 folds.
* $p < 10^{-2}$

N = 276 subjects with schizophrenia and 330 controls gathered from 5 separate sites.

Evaluated cortical thickness.

Results: inter-site prediction accuracies of 72%. Predictive signature generalized to first-episode psychosis (73% accuracy).

VBM predictive signature correlated with positive and negative symptom scores.

Identifying a neuroanatomical signature of schizophrenia, reproducible across site and stages, using machine learning with structured sparsity. De Pierrefeu (2018)

Parsimony package (github)

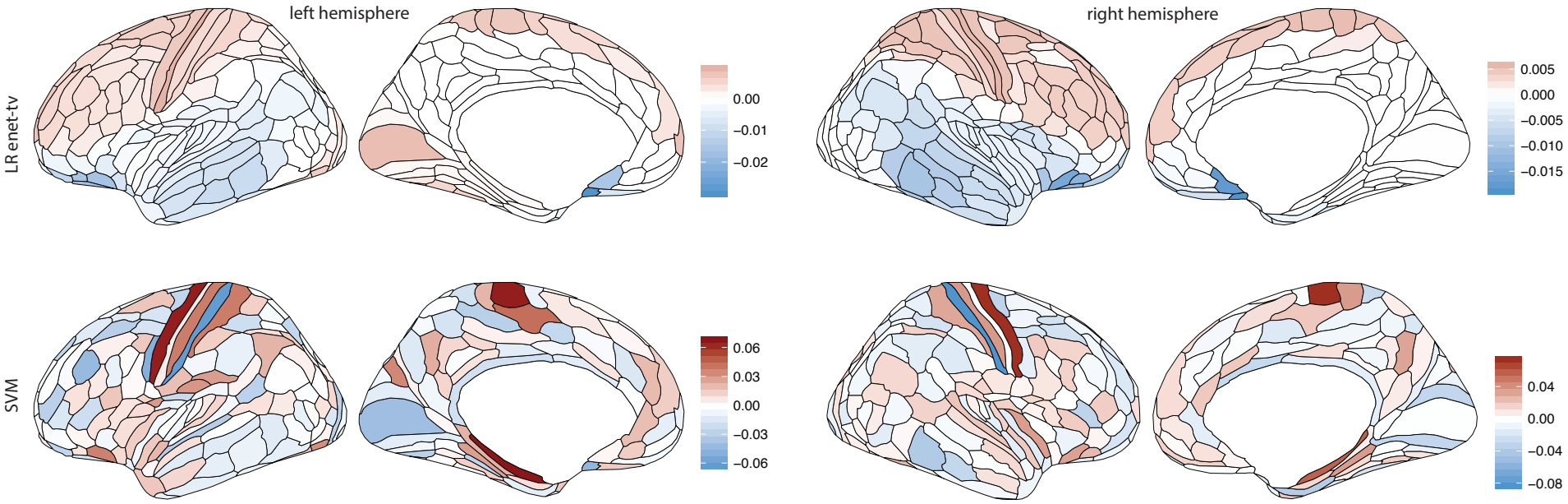
Parsimony : https://github.com/neurospin/pylearn-parsimony

- Penalties
  - L1 (Lasso)
  - L2 (Ridge)
  - Total variation (TV)
  - Overlapping group lasso (GL)
- Estimators
  - Linear regression
  - Lasso
  - Elastic Net
  - Linear Regression + L1 + L2 + TV
  - Linear Regression + L1 + L2 + GL
  - Logistic Regression + L1 + L2 + TV
  - Logistic Regression + L1 + L2 + GL
  - Linear Regression + L2 Smoothed L1 + TV

De Pierrefeu et al.,

- Machine learning with structured sparsity: application to neuroimaging-based phenotyping in autism spectrum disorder and schizophrenia. 2018. *PhD Thesis*.
- Structured sparse principal components analysis with the TV-Elastic Net penalty, *IEEE Trans Med Imaging* 2018: 37: 396 – 407
- Identifying a neuroanatomical signature of schizophrenia, reproducible across sites and stages, using machine learning with structured sparsity. *Acta Pscyhiatr Scand* 2018: 138: 571 – 580
- Prediction of activation patterns preceding hallucinations in patients with schizophrenia using machine learning with structured sparsity. *Hum Brain Mapp* 2018: 39: 1777 - 1788

Glasser parcellation as a mesh



Classifier	AUC	Precision	Recall
SVM	0.55	0.36	0.58
Enet-TV	0.59	0.65	0.58