Machine learning for Neuroimaging: an introduction

Bertrand Thirion,
INRIA Saclay-Île-de-France, Parietal team
http://parietal.saclay.inria.fr
bertrand.thirion@inria.fr

INVENTEURS DU MONDE NUMÉRIQUE

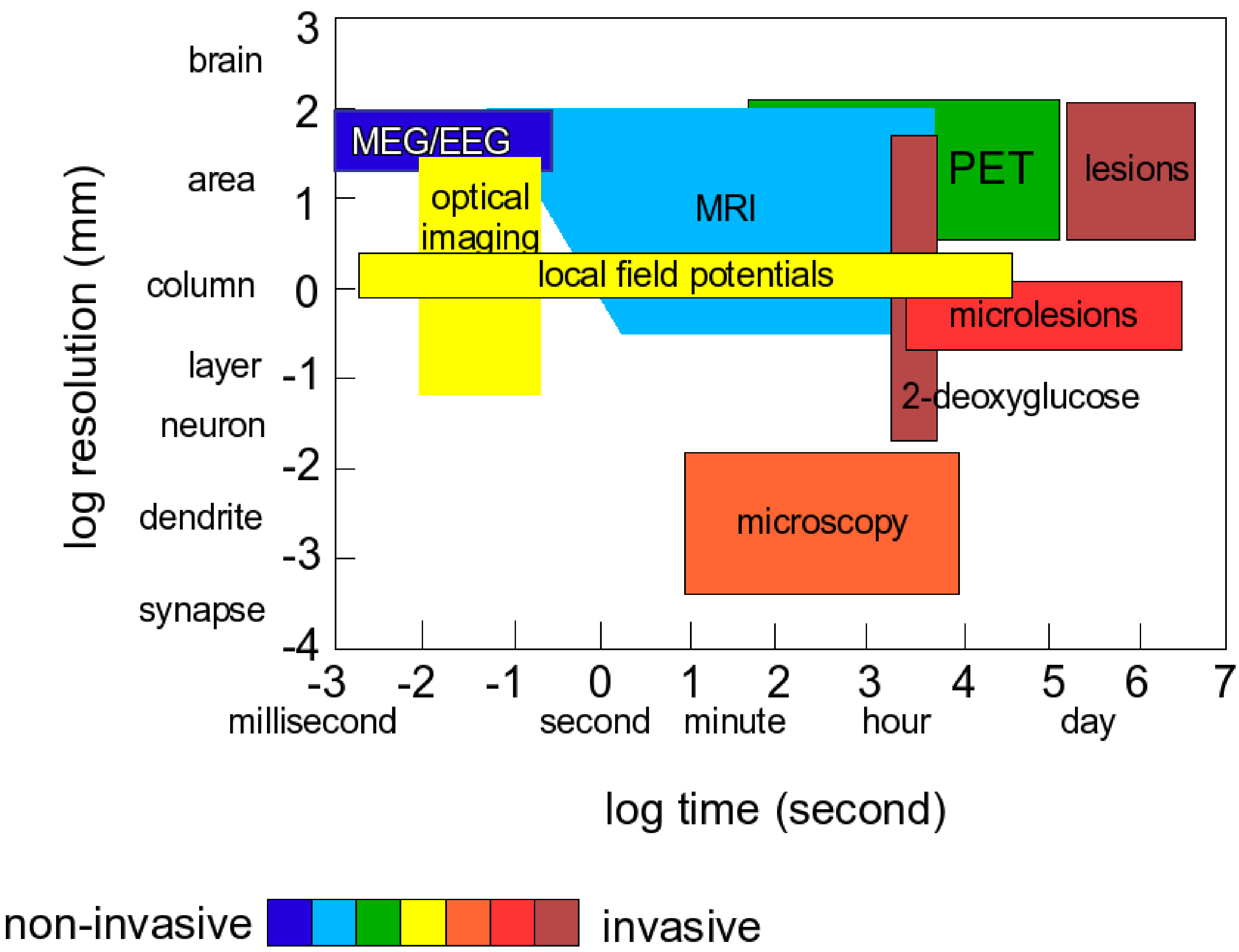
Outline

- Neuroimaging in 5'
- Some learning problems in neuroimaging:
  - Medical diagnosis & evaluation of risk factors
  - Study of between subject-variability
  - Brain reading
  - Brain connectivity mapping
- Common technical challenges

Handbook of Functional MRI Data Analysis

Russell A. Poldrack,Jeanette Mumford,Thomas Nichols

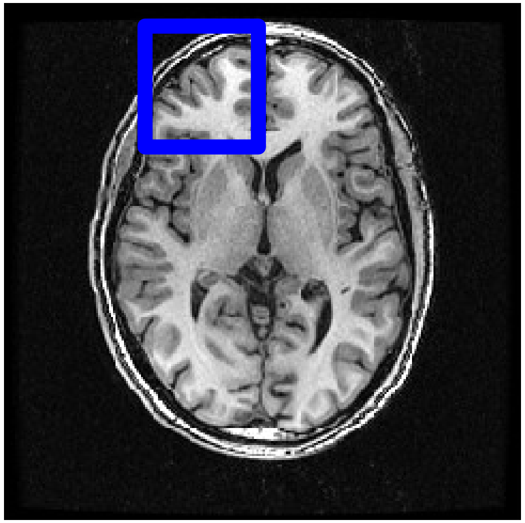
NeuroImaging: modalities and aims

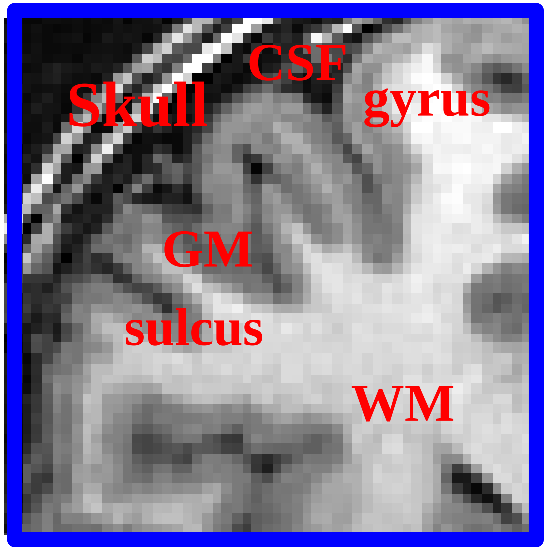


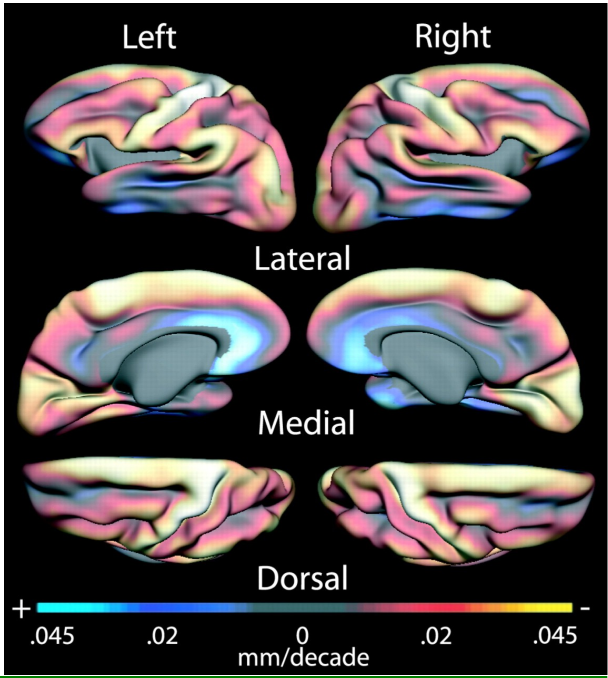
- 'Functional' (time resolved) modalities: fMRI, EEG, MEG
- vs 'anatomical' (spatially resolved) modalities: T1-MRI, DW-MRI

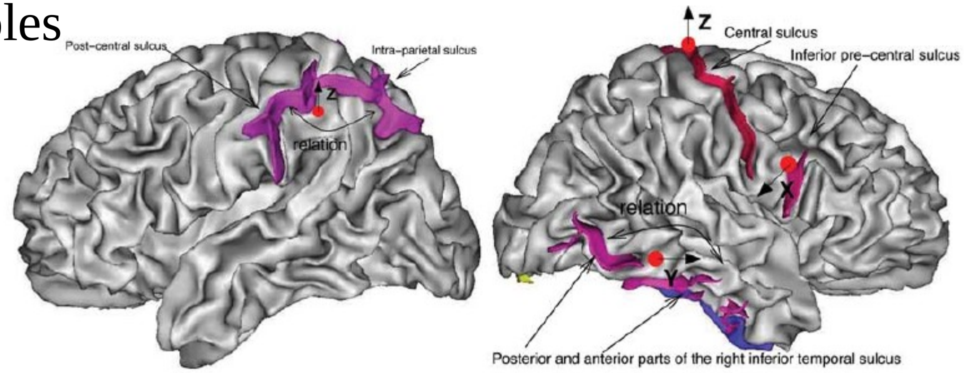
Neuroimaging modalities: T1 MRI

- T1 MRI yields
- Iconic (voxel-based) statistics
  - density of grey matter (voxel-based morphometry)
  - Cortical thickness
  - Gyrification ratio
- Landmarks-based statistics
  - Sulcus shape/orientation
- 10^2 to 10^6 variables
- $(1mm)^3$



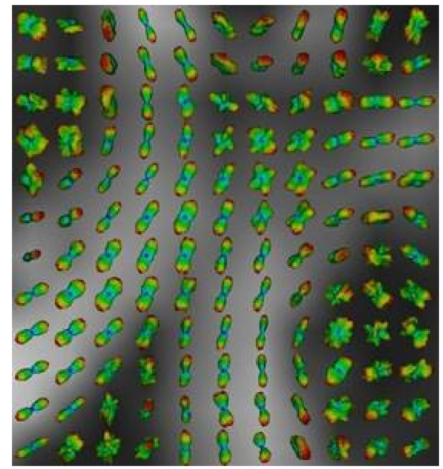


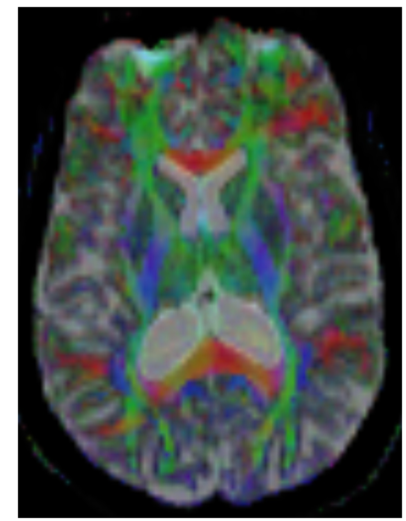


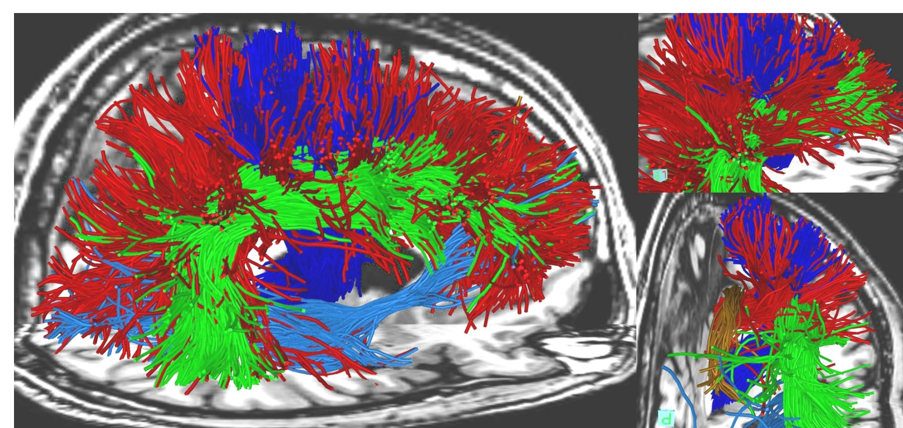


Neuroimaging modalities: DW-MRI

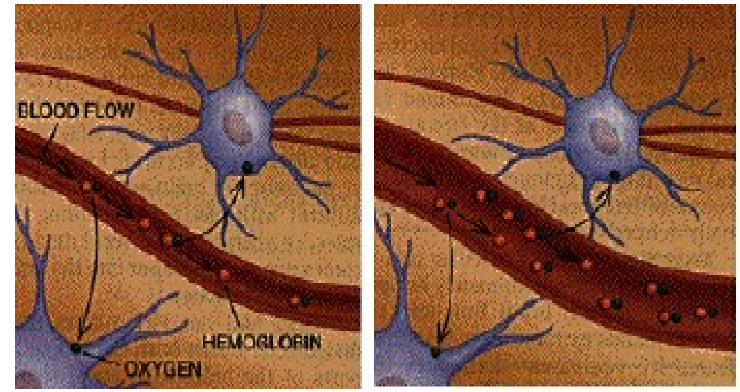
- Diffusion MRI: measurement of water diffusion in all directions in the white matter
- Resolution: $(2\text{mm})^3$, 30-60 directions
- Yields the local direction of fiber bundles that connect brain regions
- *fibers/bundles* can be reconstructed through tractography algorithms
- Statistical measurement on bundles (counting, fractional anisotropy, direction)



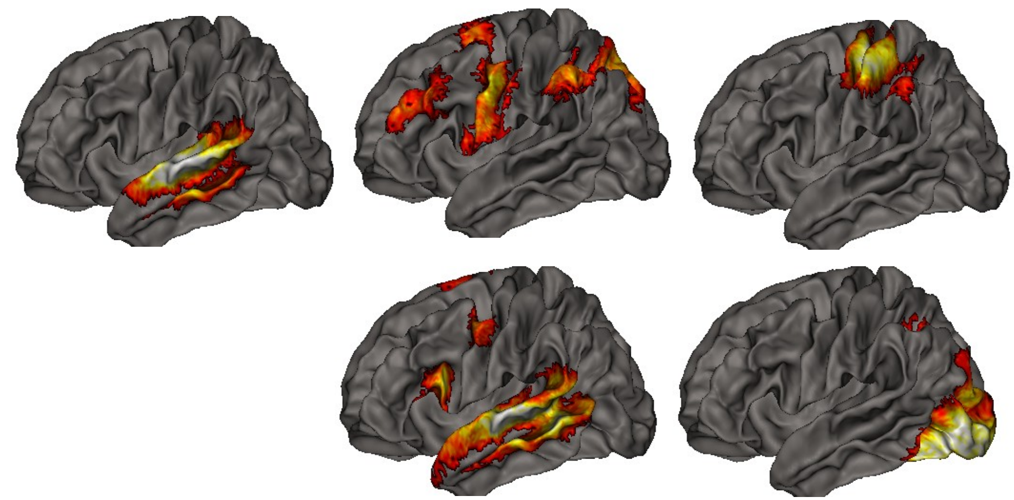




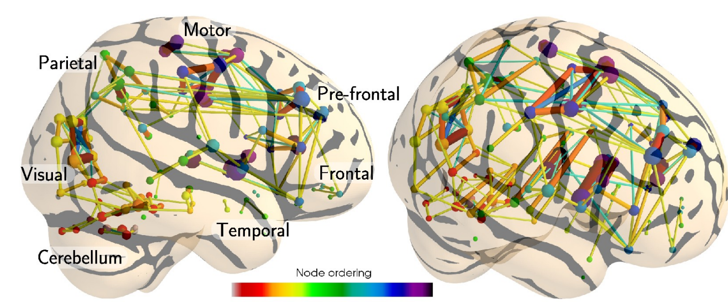
NeuroImaging modalities: fMRI



- BOLD signal: measures blood oxygenation in regions where synaptic activity occurs
  - Used to detect functionally specialized regions
  - But indirect measurement
  - Not a true quantitative measurement

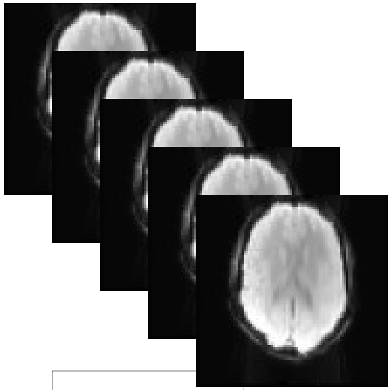


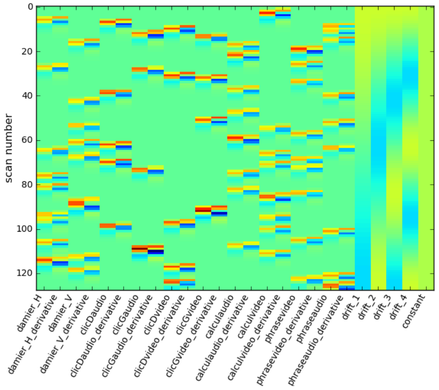
- Can also be used to characterize network structure from brain signals
- 10^2 to 10^6 observations
- Resolution $(2\text{-}3\text{mm})^3$, TR = 2-3s

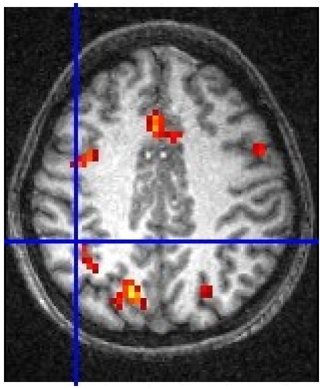


Neuroimaging data pre-processing

- Data depend on various acquisition parameters (TR, TE, resolution, FOV...)
- But also on multiple preprocessing steps,
  - which are standardized,
  - but there is room for optimization







Raw fMRI data → Space-time relaignment → Spatial normalization → GLM analysis → Activation parameters Statistical Parametric Maps

Experimental paradigm and contrasts → GLM analysis

Hemodynamic reponse function → GLM analysis

Noise model → GLM analysis

NeuroImaging: modalities and aims

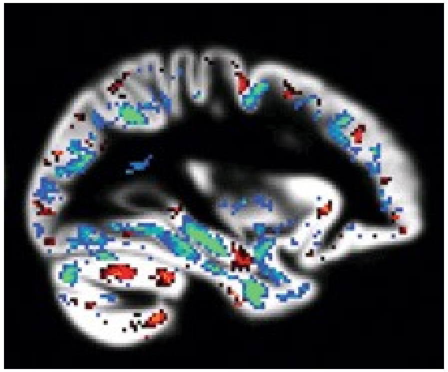
- Provide some biomarkers for diagnostic/prognostic, study of risk factors for various brain diseases
    - Psychiatric diseases
    - Neuro-degenerative diseases,
    - Brain lesions (strokes...)
- Understand brain organization and related factors: brain mapping, connectivity, architecture, development, aging, relation to behavior, relation to genetics
- Study chronometry of brain processes (MEG)
- Build brain computer interfaces (EEG)

Outline

- Neuroimaging in 5'
- Some learning problems in neuroimaging:
  - Medical diagnosis & evaluation of risk factors
  - Study of between subject-variability
  - Brain reading
  - Brain connectivity mapping
- Common technical challenges

learning in Neuroimaging: Medical diagnosis and evaluation of risk factors

- Rationale: brain images provide quantitative measurements of brain organization that reflect brain disease, abnormalities etc.
    - Cortical thickness (T1-MRI)
    - Brain shape/folding (T1-MRI)
    - Brain anatomical connections (DW-MRI)
    - Neural activation (BOLD)
    - Vascular structure/density (BOLD, ASL)



Alzheimer in VBM, [Klöppel et al., *Brain 2008]*

- Different approaches for population comparison:
classical statistics, population discrimination, outlier detection

Diagnosis based on medical images

Neuroimaging data

X-MRI, PET,...

Brain descriptors

- local (gyrification ratio on anatomical image)
- global (functional integration of brain systems)
- more meaningful than raw data + denoising

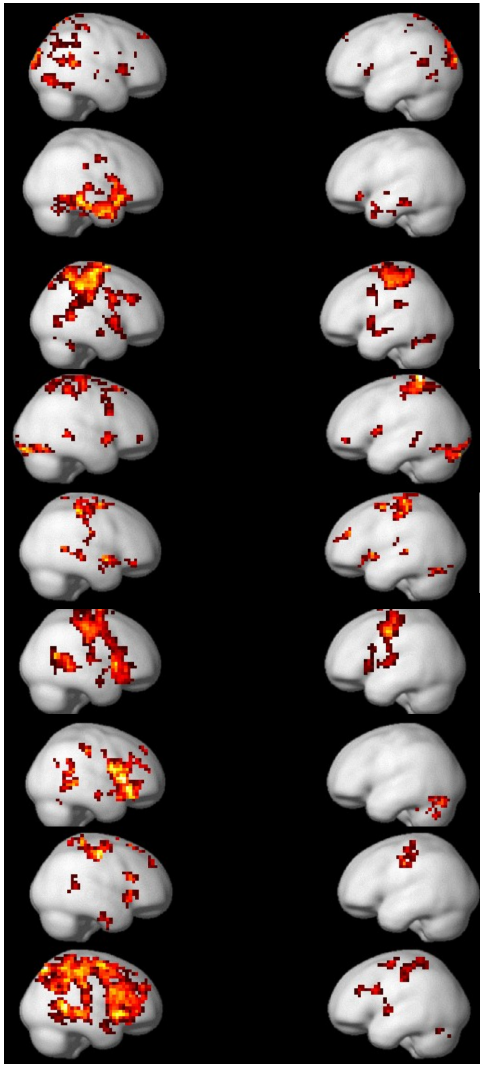
Classification/ Regression / outlier detection

Accuracy of the prediction ?
Discriminating information ?

Fundamental difficulty: Necessity to coregister brain anatomically but risk of masking brain shape differences

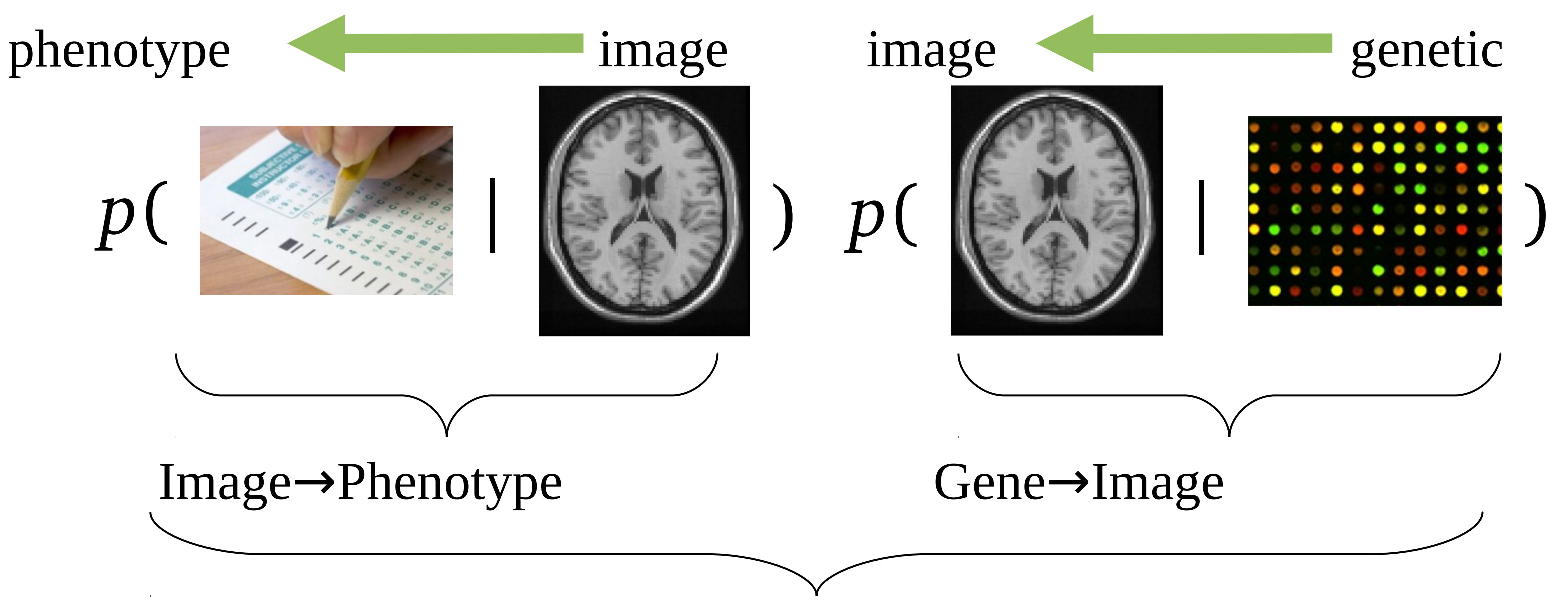
Study of between-subject variability

- Brain diseases are extreme case of *normal variability*
- Between-subject variability is a **prominent effect** in neuroimaging:
  - hard to characterize as such
  - how much of it can be explained using other data ?
- Data easier to acquire on *normal* populations
  - Confrontation to behavioral data
  - Confrontation to genetic data
- Perspective of individualized treatments



Study of between-subject variability

- The major challenge here is to discover statistical associations between complex, high-dimensional variables (regression)
- Frequently handled as unsupervised problems: describe the density of the data based on observations (manifold learning, mixture modeling)



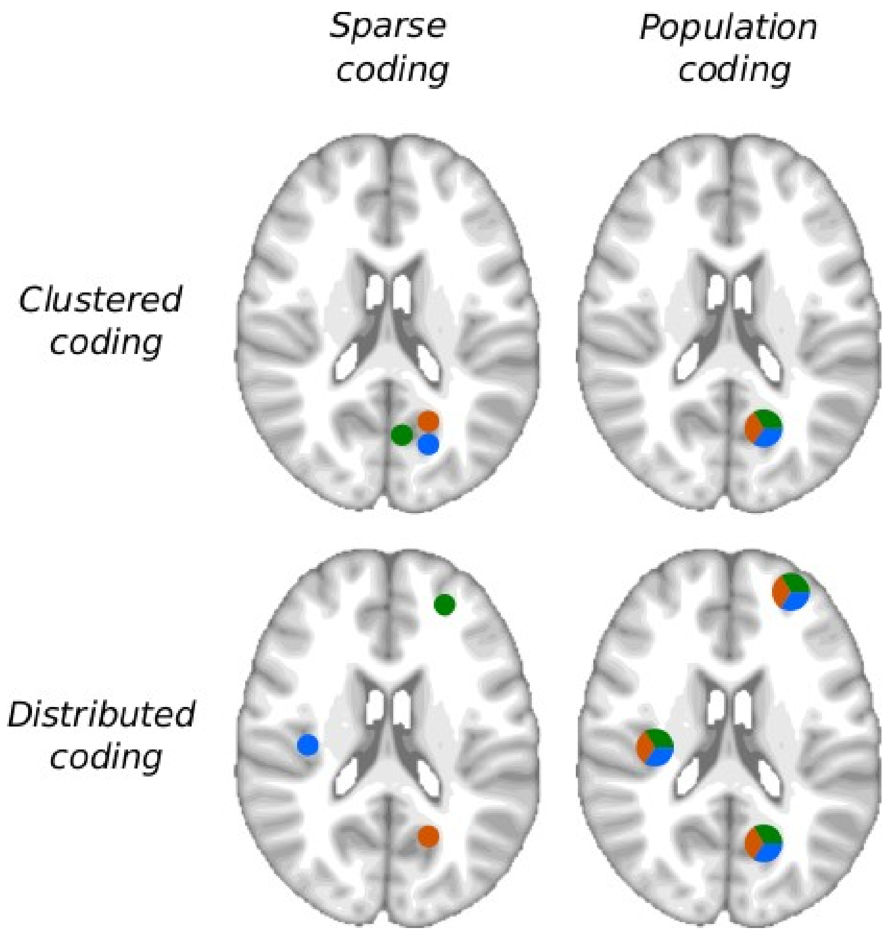
Imaging as an **intermediate (endo)phenotype**

“Brain reading”

- Definition: Use of functional neuroimaging data to infer the subject's behaviour – typically the brain response related to a certain stimulus
- Similar to BCI -to some extent-
  - without time constraints
  - More emphasis on model correctness
- Popular due to its sensitivity to detect small-amplitude but distributed brain responses
- Rationale: population coding

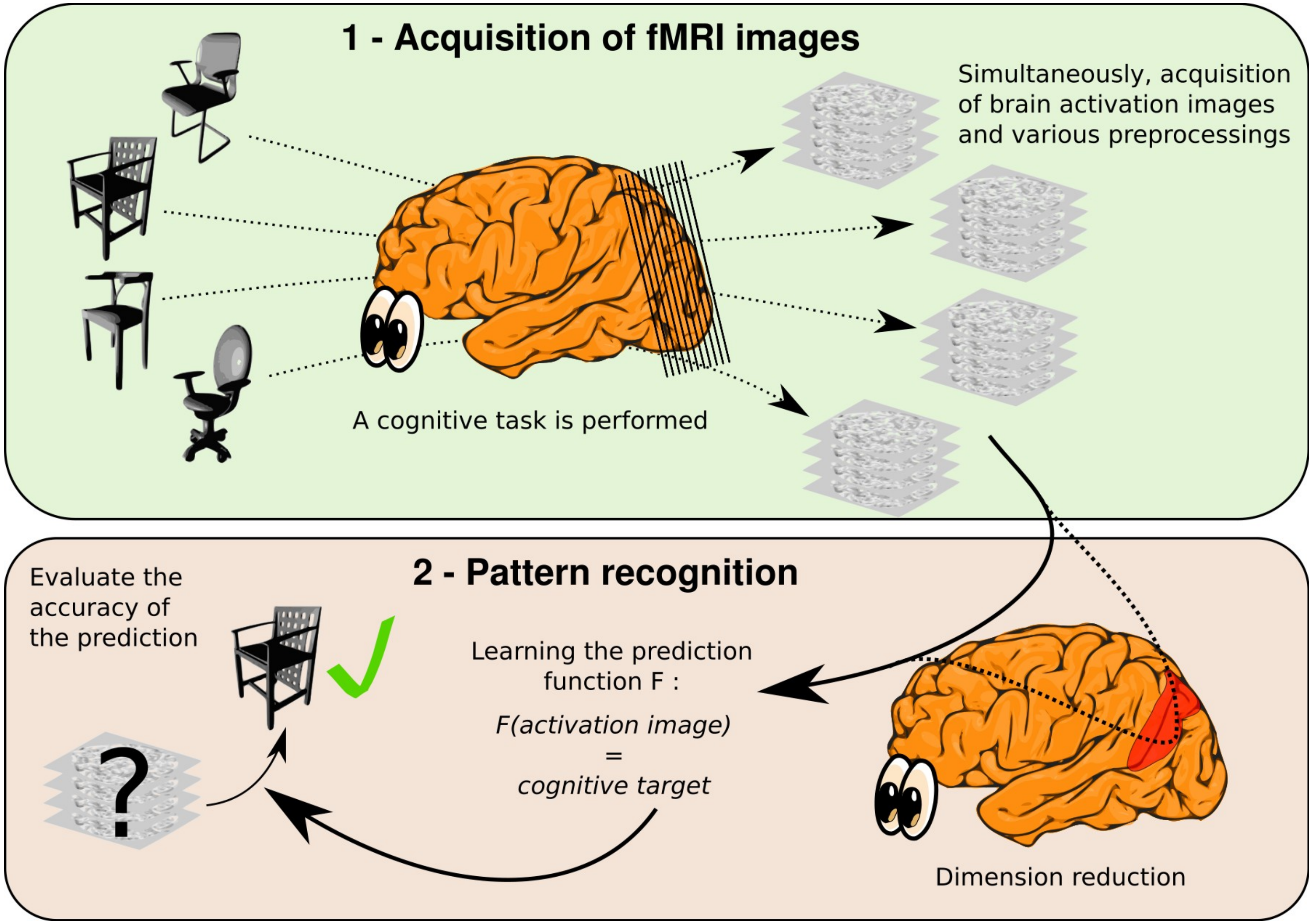
Brain reading: population coding

Different spatial models of the functional organization of neural networks



- Not a unique kind of pattern for the spatial organization of the neural code.
- This is further confounded by between-subject variability

Brain reading / Reverse inference



Aims at predicting a cognitive variable → decoding brain activity
[Dehaene et al. 1998, Cox et al. 2003]

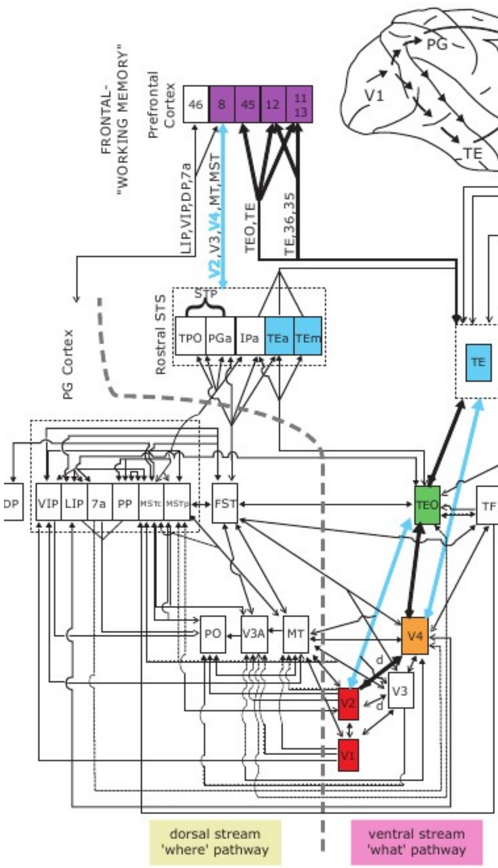
Brain reading / open issues

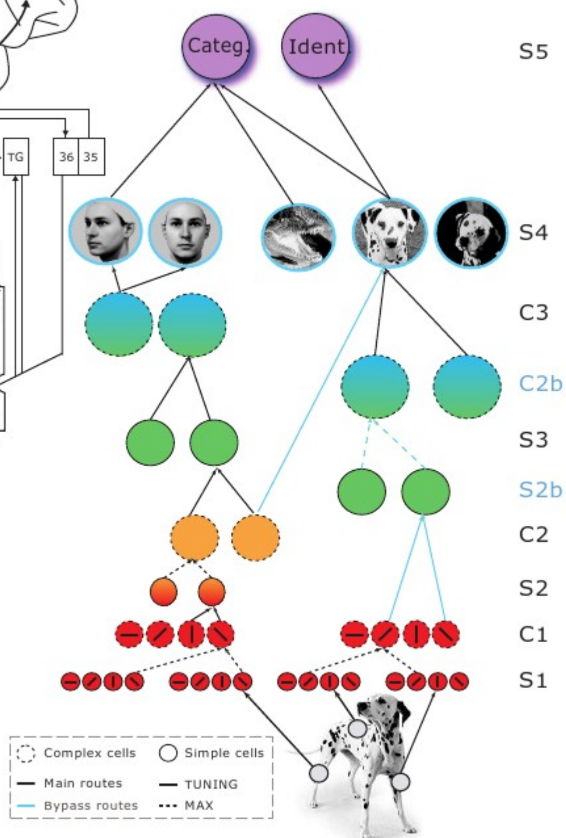
Do we want this....

```
### Compute the prediction accuracy for the different folds (i.e. session)
cv_scores = cross_val_score(anova_svc, X, y, cv=cv, n_jobs=-1,
                            verbose=1, iid=True)

### Return the corresponding mean prediction accuracy
classification_accuracy = np.sum(cv_scores) / float(n_samples)
```

```
>>> print "Classification accuracy: %f" % classification_accuracy, \
    " / Chance level: %f" % (1. / n_conditions)
Classification accuracy: 0.744213  / Chance level: 0.125000
```

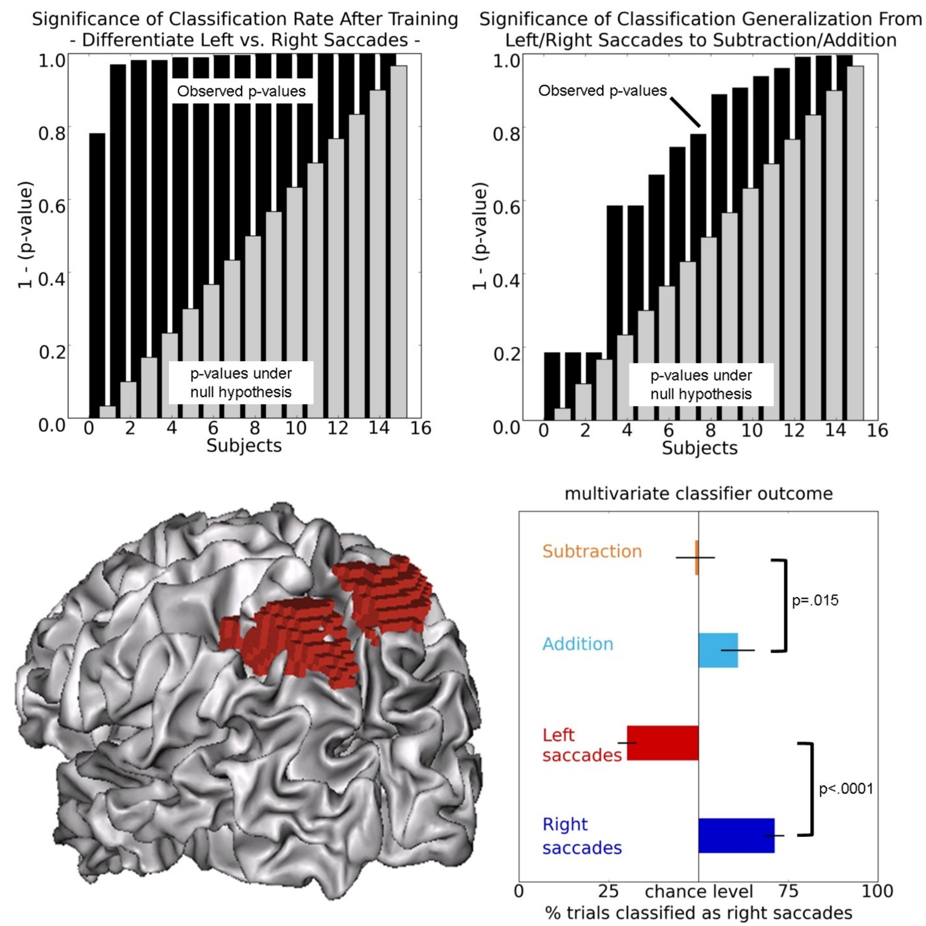




… or that ?

Brain reading/open issues

- a classifier trained to discriminate left versus right saccades can also *decode* mental arithmetics:
- left saccade ⇔ subtraction
- right saccade ⇔ addition
- This generalization occurs only when based on two regions of the parietal cortex
- This shows that the same neural populations are involved in ocular saccades and arithmetics



[Knops et al., *science* 2009]

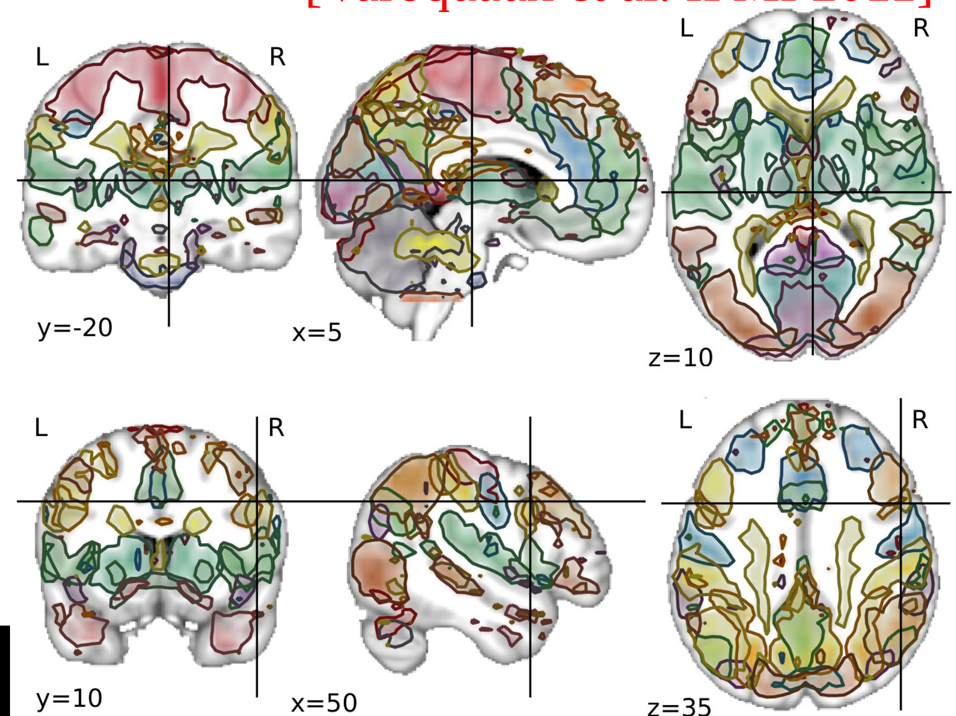
Functional connectivity mapping

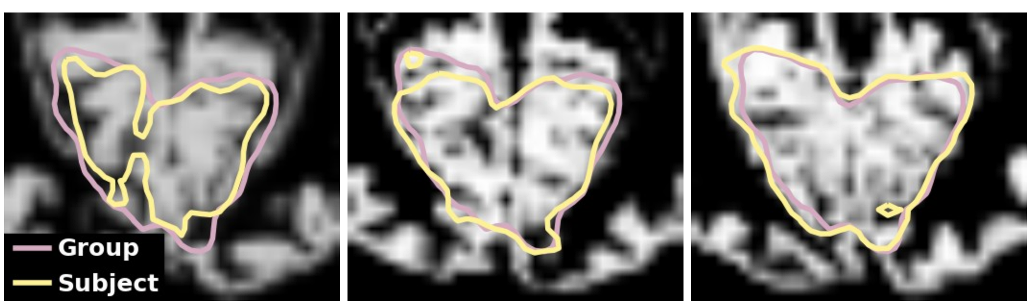
- Definition: consists in deriving a quantitative measure of brain networks integration based on functional neuroimaging observation
- Rationale
  - Popularity of resting-state fMRI.
  - Model-driven approach (SEM, DCM): restrictive hypotheses
- Learning problems
  - Segment regions based on connectivity information (common to many neuroimaging problems)
  - Inference of connectivity models

Learning in FCM (1)

[Varoquaux et al. IPMI 2011]

- Learn a spatial model (atlas) from the resting state data
  - ICA, clustering provide little guarantees on the result
  - Dictionary learning can be used instead



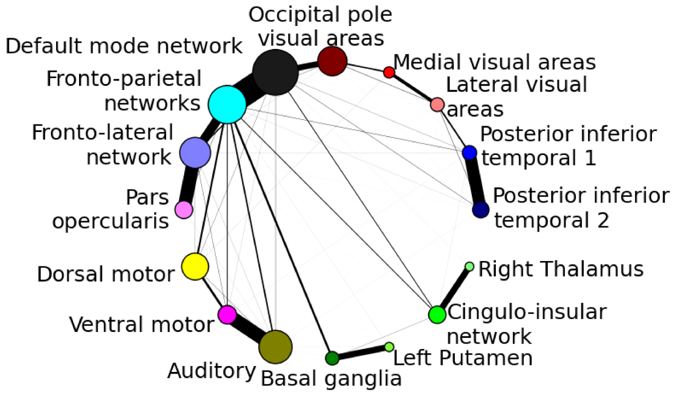


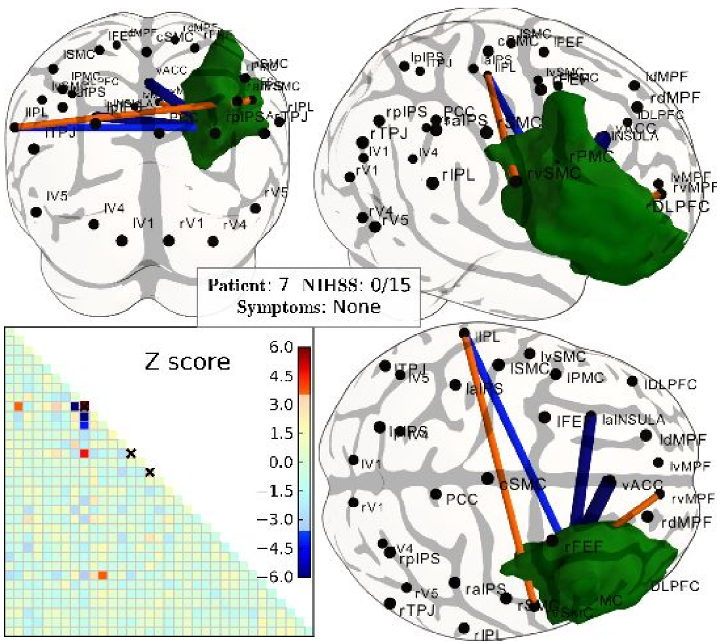
The population-level model adapts to individual configurations

Note: Atlas learning is not tied to Functional Connectivity Mapping, but is important in different contexts (parcellations)

Learning in FCM (2)

- Next: Given a set of regions, quantify properly their interactions/integration of the underlying networks
    - Threshold correlation graph → graph statistics, graph embedding
    - Learn covariance model between the set of regions (partial correlations)
- Do statistical inference on these objects





$$\left(\hat{\mathbf{K}}^{(s)}_{\ell_{21}}\right)_{s=1..S} = \operatorname{argmin}_{\mathbf{K}^{(s)}\succ 0}\left(\sum_{s=1}^{S}\left(\operatorname{tr}(\mathbf{K}^{(s)}\,\hat{\mathbf{\Sigma}}^{(s)}_{\text{sample}}) - \log\det \mathbf{K}^{(s)}\right) + \lambda \sum_{i\neq j}\|\mathbf{K}^{(\cdot)}_{ij}\|_2\right)$$

Outline

- Neuroimaging in 5'
- Some learning problems in neuroimaging:
  - Medical diagnosis & evaluation of risk factors
  - Study of between subject-variability
  - Brain reading
  - Brain connectivity mapping
- Common technical challenges

Technical challenges in MLNI

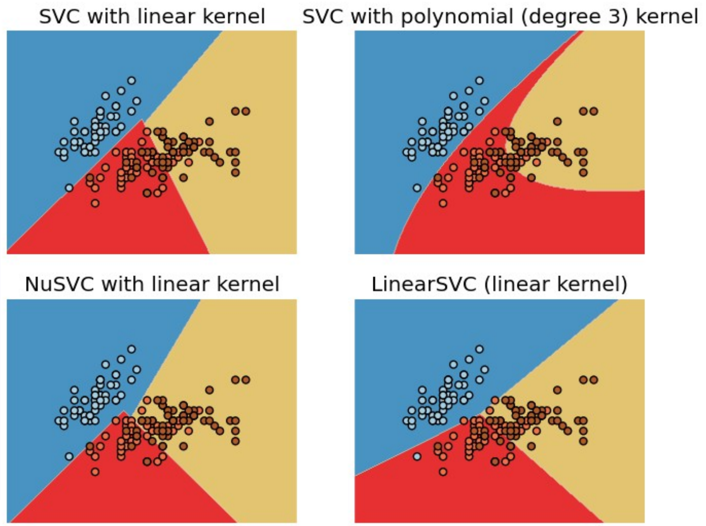
- Low SNR in the data
  - Only a fraction of the data is modeled (BOLD)
  - Presence of structured noise (noise is not i.i.d. Gaussian !) + non-stationarity in time and space
  - Few salient structures (resting-state fMRI...)
- Size of the data
  - 10^4 to 10^6 voxels in most settings
  - Compared to 10 to 10^2 samples available
- Related to the particular learning problems

Technical challenges in MLNI

- *Diagnostic/classification* problems
  - Needs accuracy mostly (+ robustness)
  - Suffers from curse of dimensionality, but this is well addressed in the literature

    [generic approaches perform well]
  - But: not the main aim of most neuroimaging studies

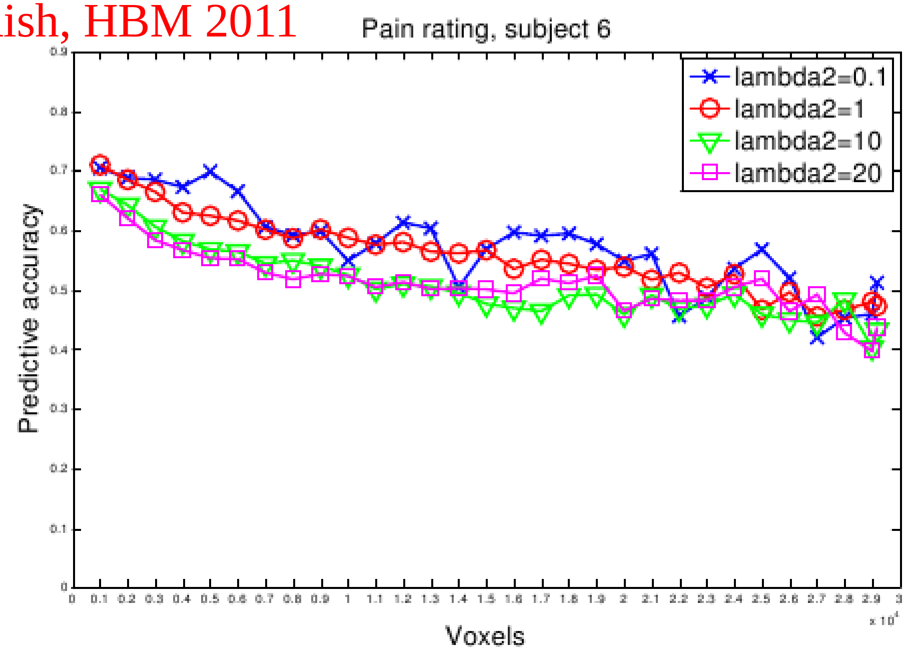
- Need a large set of tools to be compared against each other
- Need to take into account some priors on the data/true model (smoothness, sparsity)



Technical challenges in MLNI

- Recovery: retrieve the true model that accounts for the data
    - This is the main topic of all neuroimaging / brain mapping / decoding literature.
    - Suffers much more from feature dimensionality and correlation
    - Virtually in-addressed/unseen so far

I. Rish, HBM 2011



1. learn EN model for pain perception rating using first 120 TRs for training and next 120 TRs for testing.
2. Find ‘best-predicting’ 1000 voxels using EN, delete them, find next 1000 best-predicting, etc.

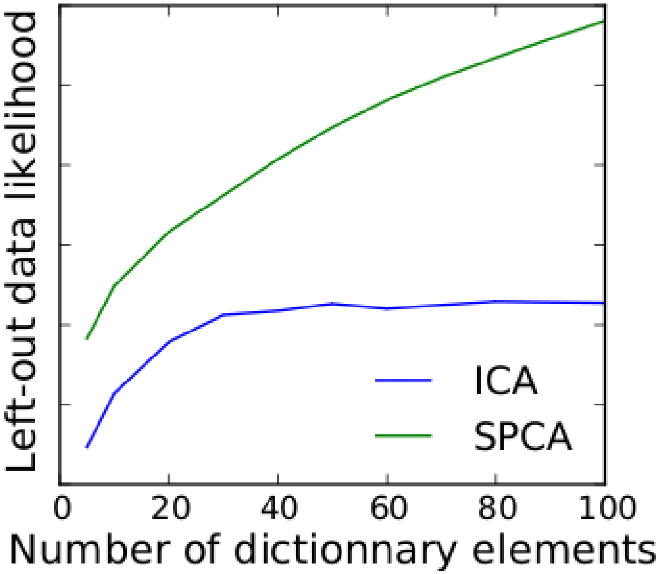
Does the predictive accuracy degrade sharply?

Surprisingly, the answer is ‘NO’

Technical challenges in MLNI

- **Unsupervised problems**: find the right model that accounts for the data, without making in prediction (number PCA components, clusters etc.)
    - Suffers from data correlation, lack of salient structure,
    - Very difficult (no happy curve around)

Likelihood of left out data in a 3-fold stratified cross-validation on resting-state fMRI data



This is not a conclusion

- To me Neuroimaging methodologists need
  - Implementations of various ML tools
    - Only the easily available tools are used; this is part of the success of SVM
    - Open-source, tested, documented ;-)
  - Guidelines on cross-validation (people tend to use leave-one-out without further thinking)
  - Steady warning against the danger of overfitting (the more complex the method, the most likely overfit occurs)
    - Working on only one dataset is one type of over-fitting
  - Understand the limitations of current methods regarding recovery and unsupervised problems.