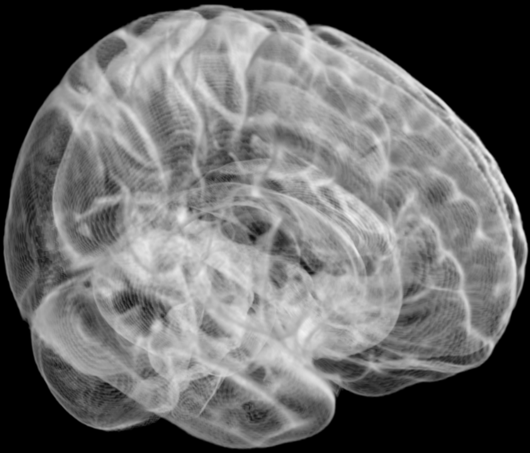
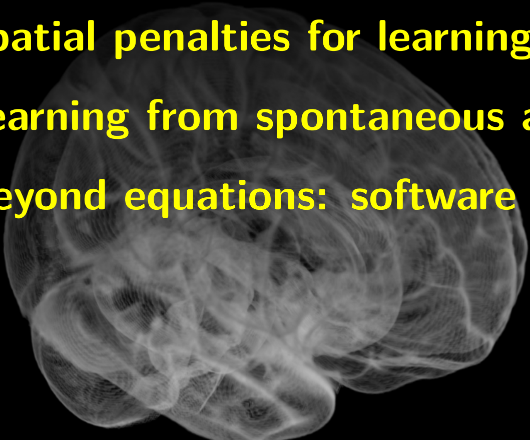
Machine learning for brain imaging

Gaël Varoquaux

INRIA/Parietal



1. An introduction to brain imaging
2. Learning to diagnose
3. Understanding brain function
4. Spatial penalties for learning from images
5. Learning from spontaneous activity
6. Beyond equations: software

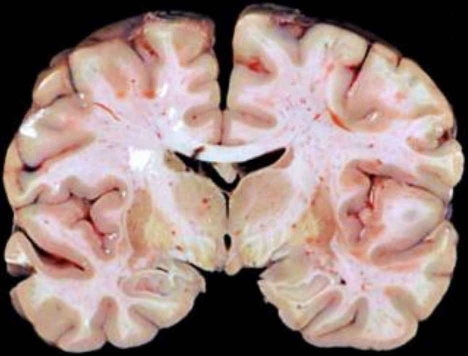


1 An introduction to brain imaging

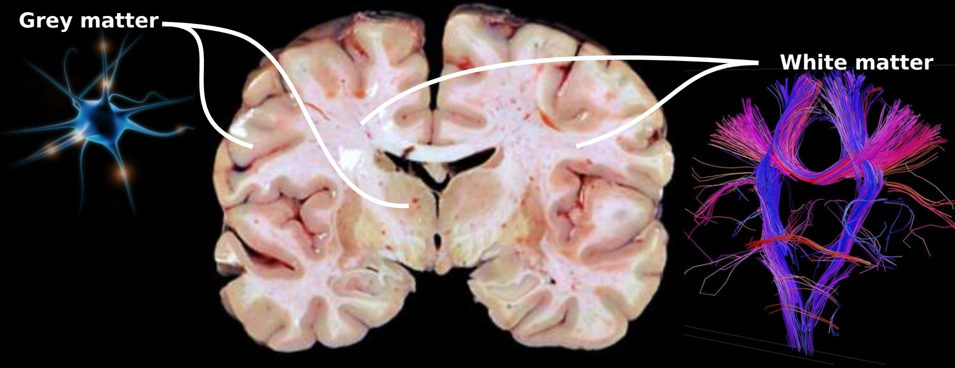
- The brain: its anatomy and function
- Imaging the brain

The brain: its anatomy and function

- Two hemispheres
- “sucli”

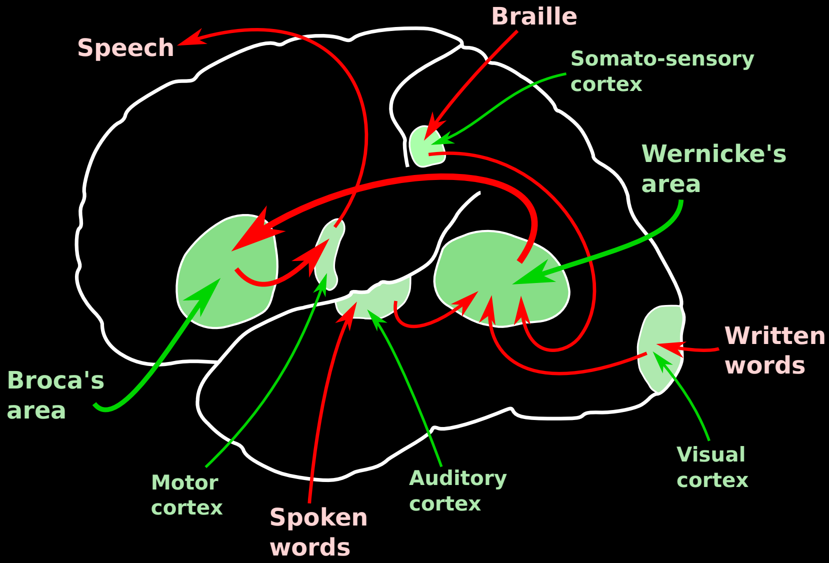


1 Inside a brain: neuroanatomy in one minute

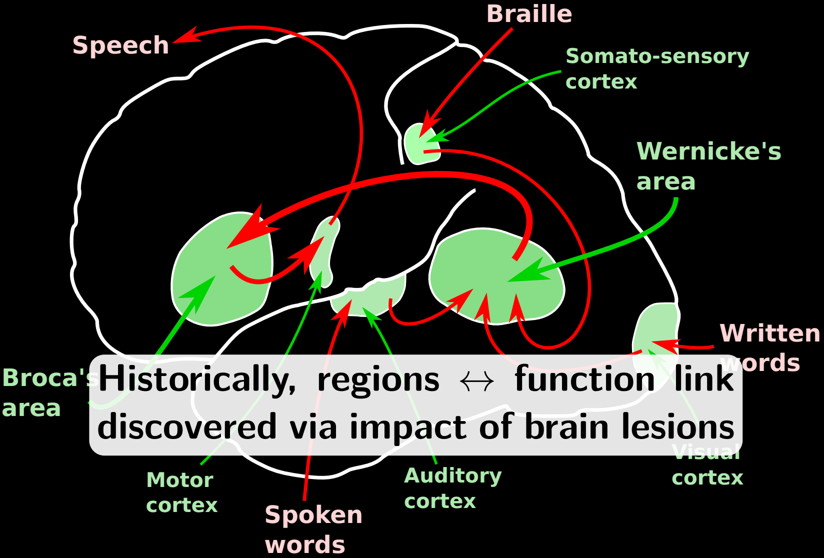


1 Inside a brain: neuroanatomy in one minute

The language circuit:

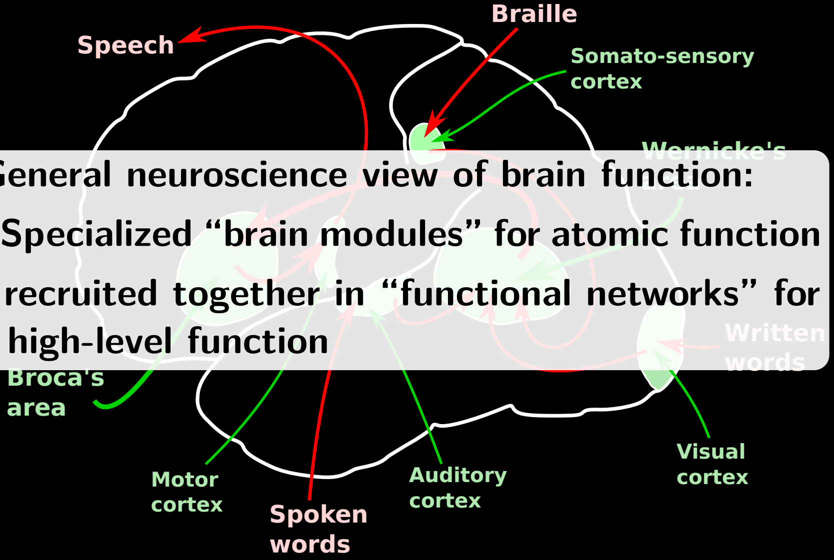


The language circuit:



Historically, regions $\leftrightarrow$ function link discovered via impact of brain lesions

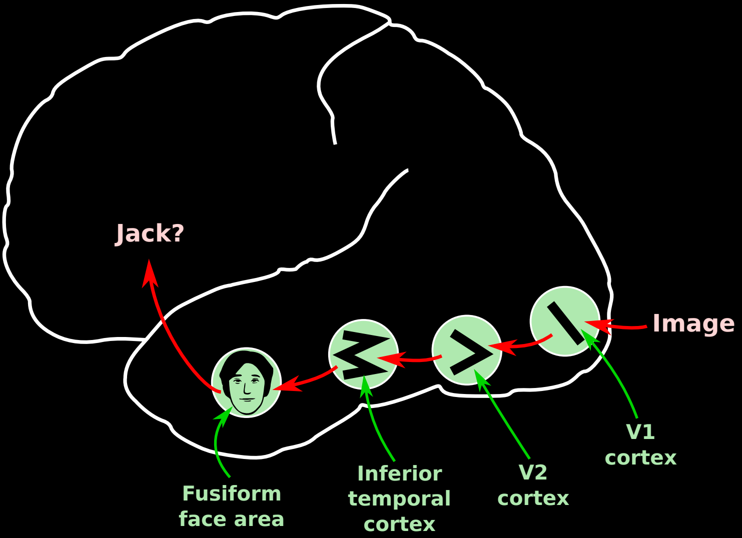
The language circuit:



General neuroscience view of brain function:

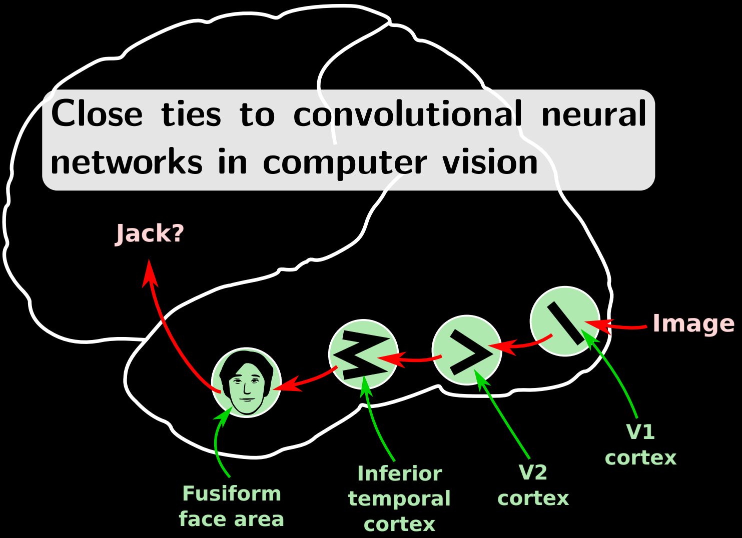
- Specialized “brain modules” for atomic function
- recruited together in “functional networks” for high-level function

The visual system: a computational model

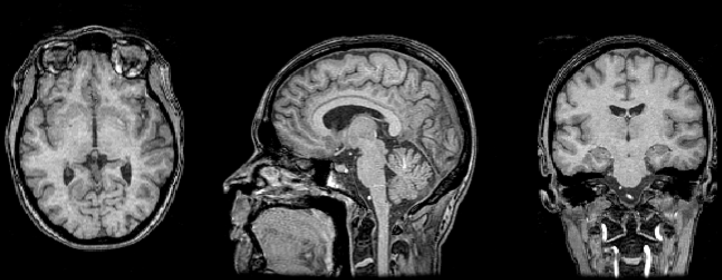


The visual system: a computational model

Close ties to convolutional neural networks in computer vision

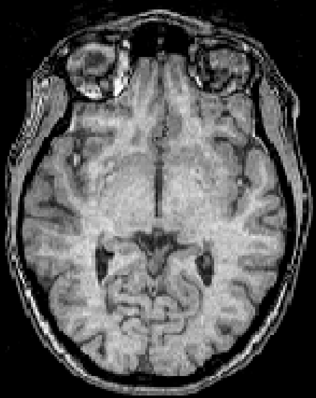


Imaging the brain



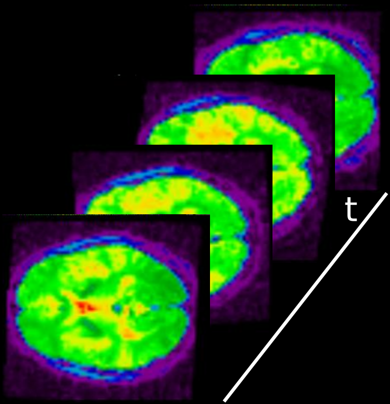
3D images

1 anatomical MRI

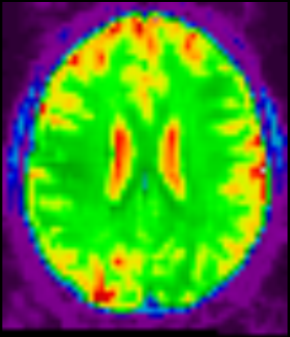


- White matter, grey matter
- Cortex

1 functional MRI (fMRI)

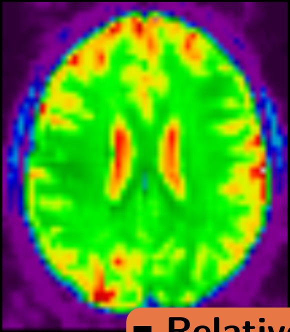


Time-resolved recordings of brain activity



Blood Oxygen Level Dependent effect

- MRI probes local magnetic properties
- oxyHemoglobin and deoxyHemoglobin have different magnetic susceptibility
- Neural activity consumes oxygen
  $\Rightarrow$ Initial dip in oxyHemoglobin
- Metabolism compensates
  $\Rightarrow$ Increase in oxyHemoglobin



Blood Oxygen Level Dependent effect

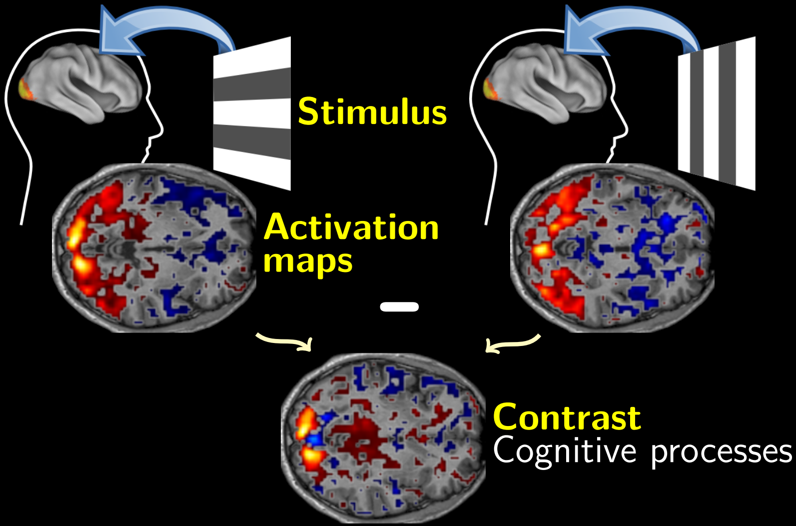
- MRI probes local magnetic properties
- oxyHemoglobin and deoxyHemoglobin have different

- Neural
  ⇒
- Metab
  ⇒

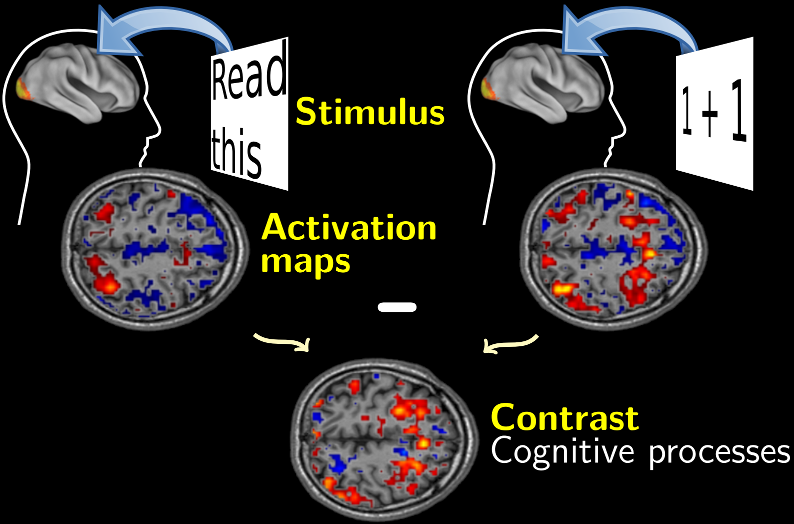
- **Relative effect, not absolute**
  ⇒ Need to contrast values
- **Very indirect effect**
  ⇒ Non-linearities, inhomogeneities, lags...

All models are wrong

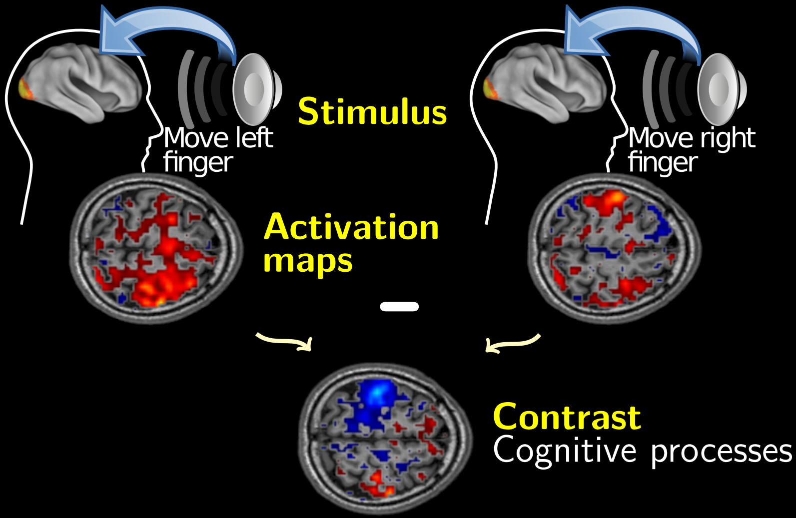
1 Mapping cognition with fMRI



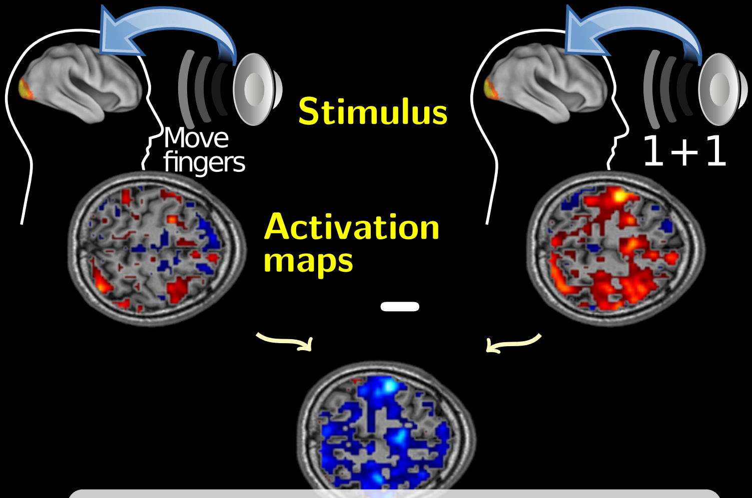
1 Mapping cognition with fMRI



1 Mapping cognition with fMRI



1 Mapping cognition with fMRI

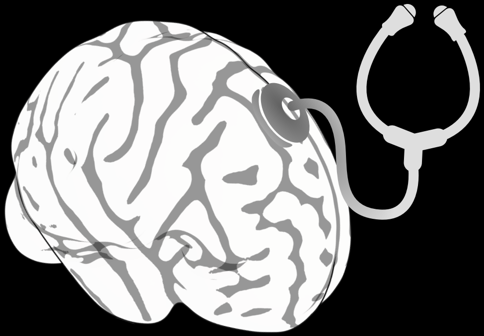


Careful crafting of contrasts to isolate high-level cognition

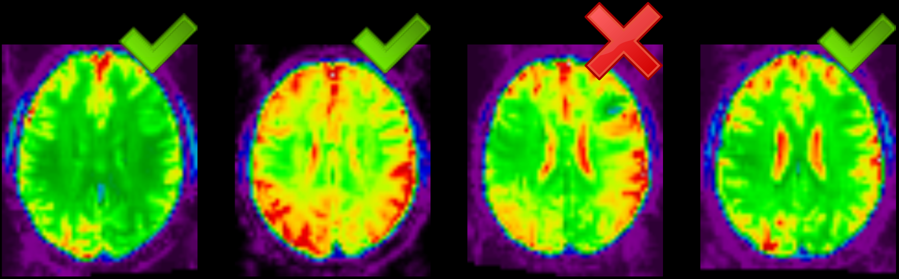
1 Magneto and electro encephalography

- Measure electromagnetic field created by neural fixing
- "Interesting" inverse problem to reconstruct sources on cortex
  $\Rightarrow$ Poor spatial resolution
- Great temporal resolution

2 Learning to diagnose



Diagnostic applications: promises and challenges



Diagnosis

Finding the nature or cause of a disease condition

Pronosis

Predicting the future evolution of the condition

$\Rightarrow$ Therapeutic indications

Early biomarkers

Measures enabling the detection of disease before standard symptoms

$\Rightarrow$ Population screening if cheap

Quantitative biomarkers

Metric to follow disease progression

$\Rightarrow$ Drug development

2 Specificity – sensitivity trade-offs

Depending on application, different types of error may need to be weighted differently

E.g.: Screening before human check
Low false negative rate, high false positive rate

Detection leading to surgery
Low false positive rate

2 More than prediction accuracy

Cannot replace the physician:

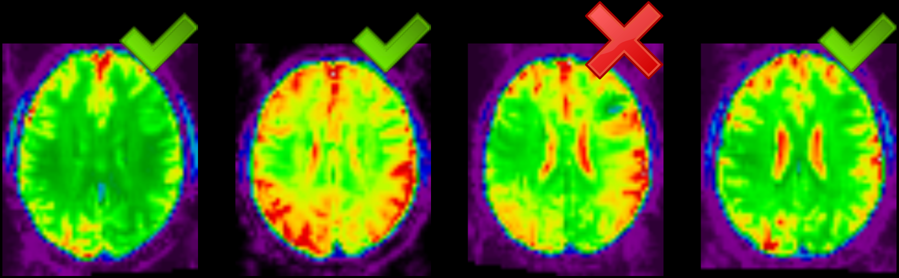
- Patient history
- Therapeutic strategies subject to logistics

...

$\Rightarrow$ No black-box

Segmentation, denoising task

as much as prediction



2 More than prediction accuracy

Cannot replace the physician:

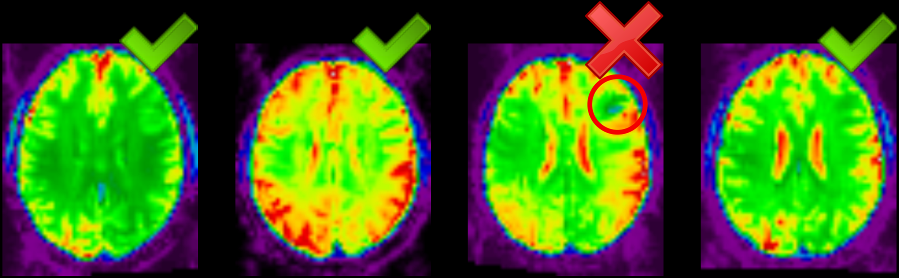
- Patient history
- Therapeutic strategies subject to logistics

  ...

⇒ No black-box

Segmentation, denoising task

as much as prediction



2 The training set is not representative

- Early or weakly-symptomatic patients not represented
  these are the most interesting
- Label noise: wrong diagnostic on difficult patients
  $\Rightarrow$ validation difficult
- Confounding factors for patients

$\Rightarrow$ Epidemiological studies (biobank, UK)
select subjects randomly from
normal population, and follow them

but imaging cost prohibitive

2 The training set is not representative

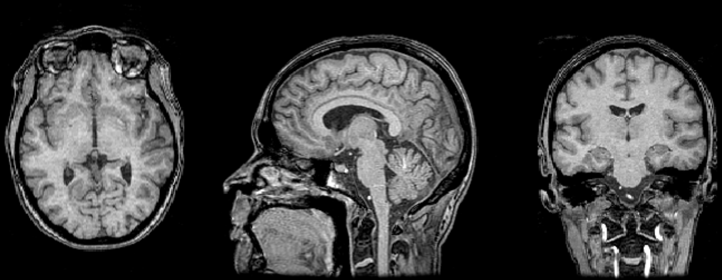
- Early or weakly-symptomatic patients not represented
  these are the most interesting
- Label noise: wrong diagnostic on difficult patients
  $\Rightarrow$ validation difficult
- Confounding factors for patients

Real problems are not patient/control classifications, but multi-pathology classifications

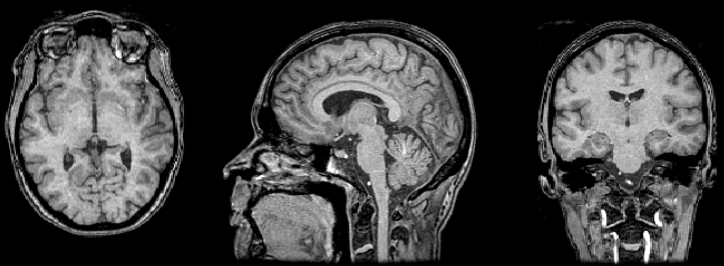
Datasets are small (not many subjects)

- **Researchers don't share data**
- **Large inter-site variability**

Features from brain images



Features from brain images



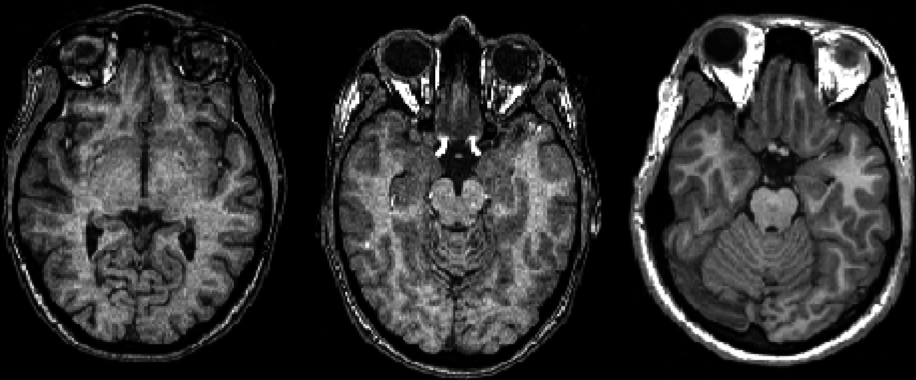
Not a standard computer vision pipeline!

- Brains are not translation-invariant
- We don't understand much about the brain

  but we still know *something*

What are we trying to capture?

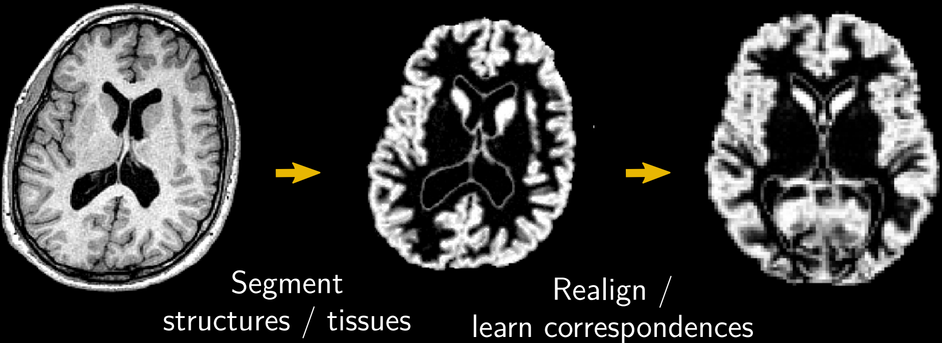
- Changes in brain shape?
- Local grey-matter changes?
- Micro-lesions?



- Changes in brain shape?
- Local grey-matter changes?
- Micro-lesions?

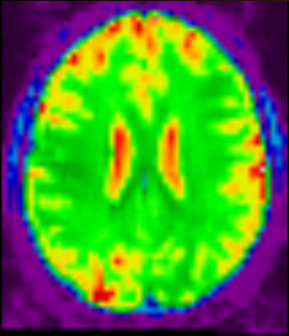
Across subjects

Feature extraction in anatomical images



Features in structures or in correspondence vectors

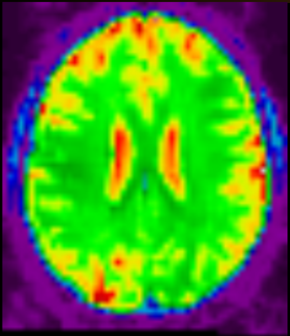
- Voxel-based morphometry: grey matter density
- Cortical thickness
- Shape descriptors of cortex
- Realignment transformation field



Functional images are low SNR

Why use them?

2 Functional images



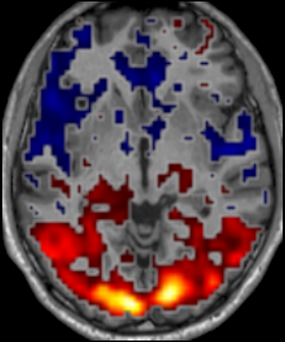
Functional images are low SNR

Why use them?

- Behavioral deficit vary
  for a given lesion
- Neuro-psychiatric diseases (autism, schizophrenia) = deficit of function with no known anatomy

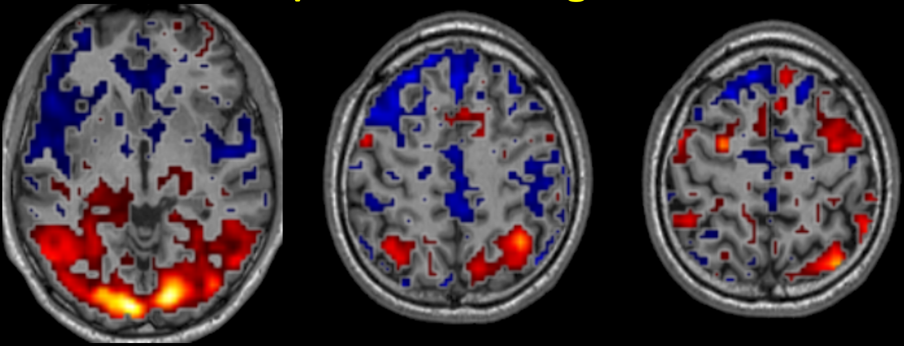
Des troubles de la parole

Activation maps to dissect cognitive effects



Images

Activation maps to dissect cognitive effects



Images Reading Counting

Often: one map per stimuli presentation

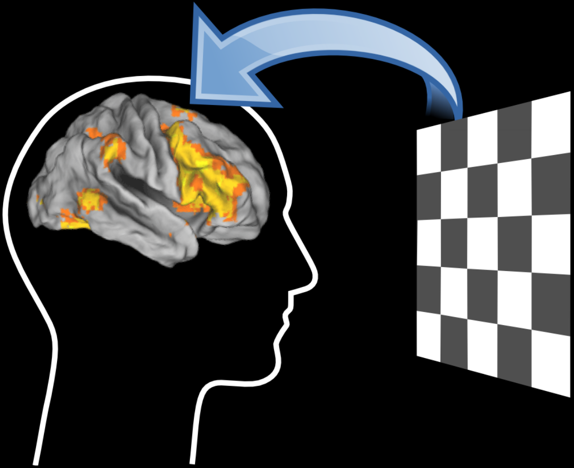
[Poldrack 2011]

3 Understanding brain function

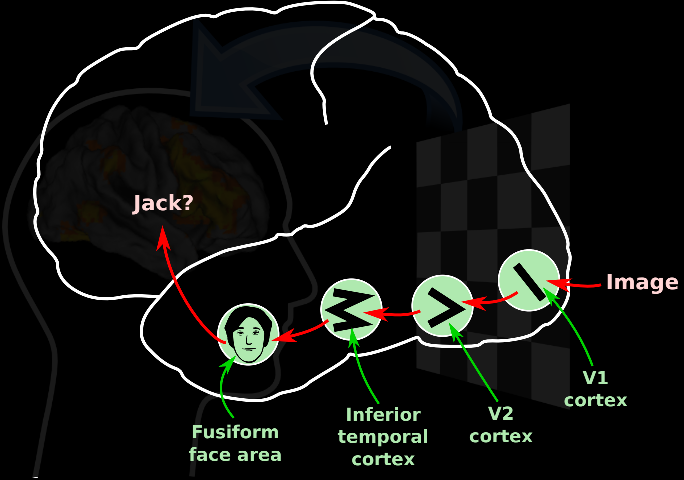
Cognitive neuroimaging

3 Machine learning for cognitive neuroImaging

Learn a bilateral link between brain activity and cognitive function

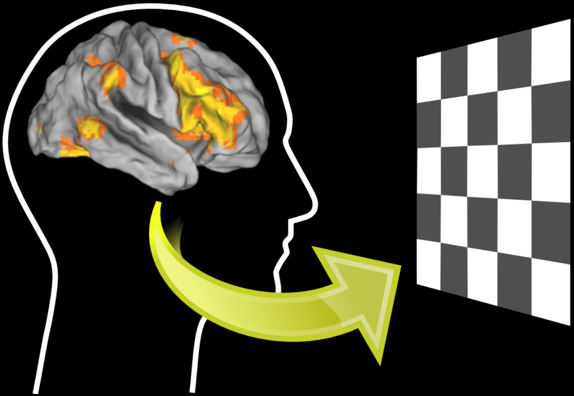


Predicting neural response:
encoding models



Predicting neural response:

encoding models

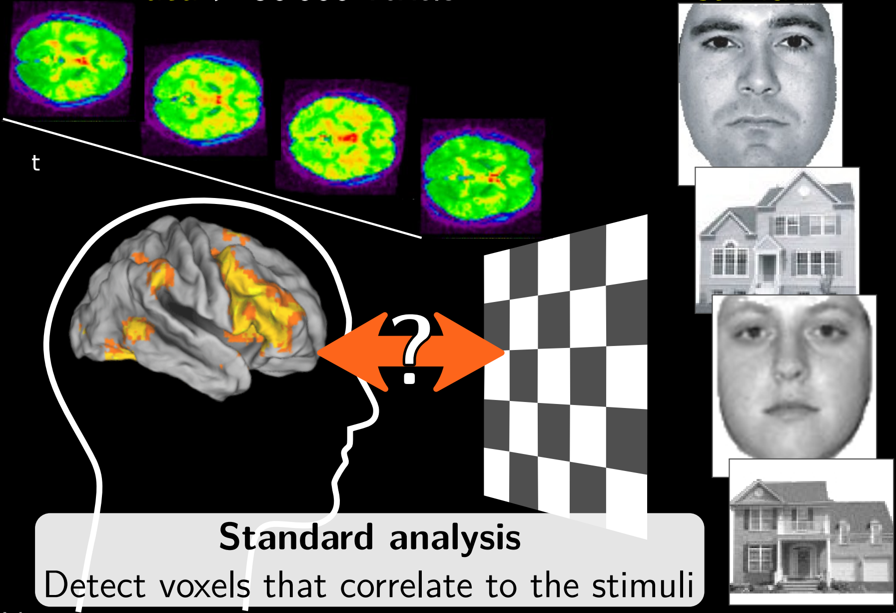


"Brain reading": **decoding**

3 Brain mapping

fMRI data > 50 000 voxels

stimuli



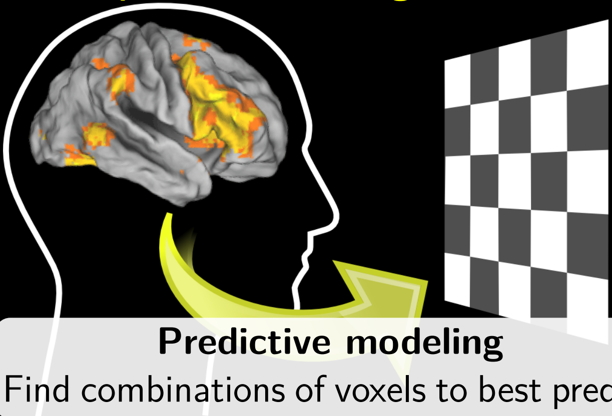
Standard analysis

Detect voxels that correlate to the stimuli

- Predicting the object category viewed

  [Haxby 2001, *Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex*]

Supervised learning task

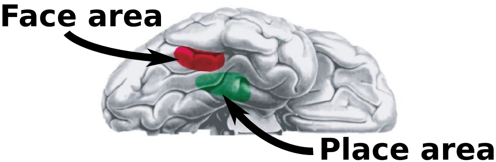


Predictive modeling

Find combinations of voxels to best predict

- Predicting the object category viewed
  [Haxby 2001, *Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex*]

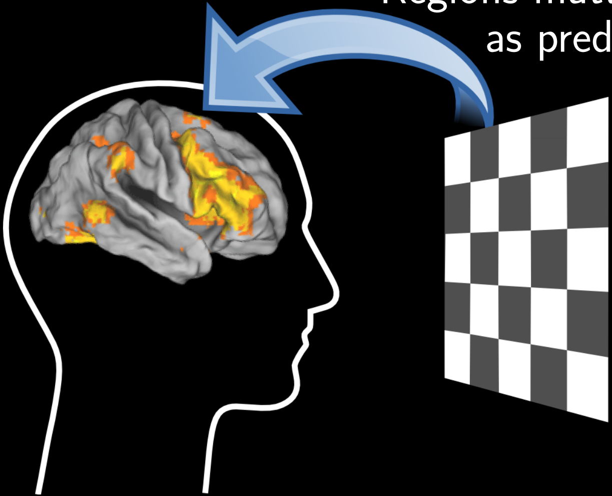
Take home message:
brain regions, not prediction



Predictive modeling
Find combinations of voxels to best predict

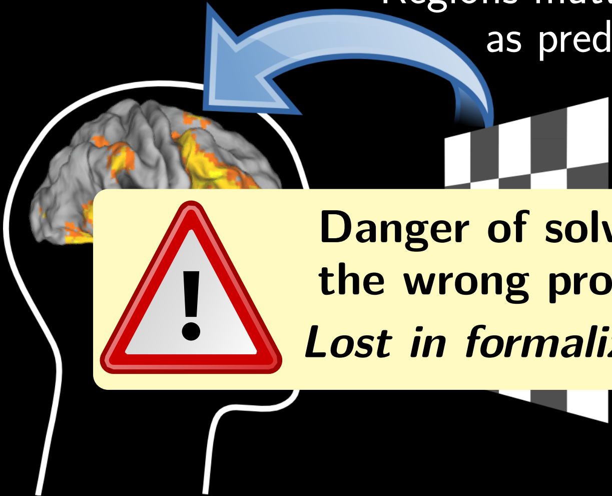
Recovery rather than prediction

Regions matter as much
as prediction score



Recovery rather than prediction

Regions matter as much
as prediction score

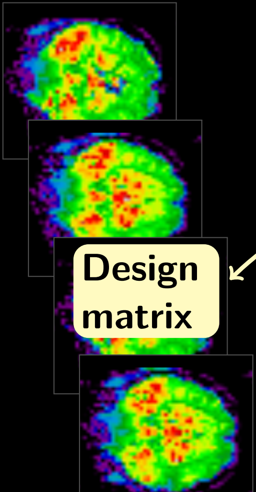


Danger of solving the wrong problem

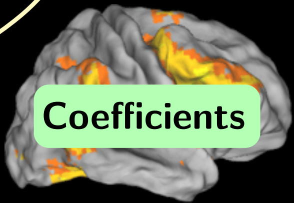
Lost in formalization

4 Spatial penalties for learning from images

$$\mathrm{sign}(\mathbf{X}\,\mathbf{w} + \mathbf{e}) = \mathbf{y}$$



$\times$



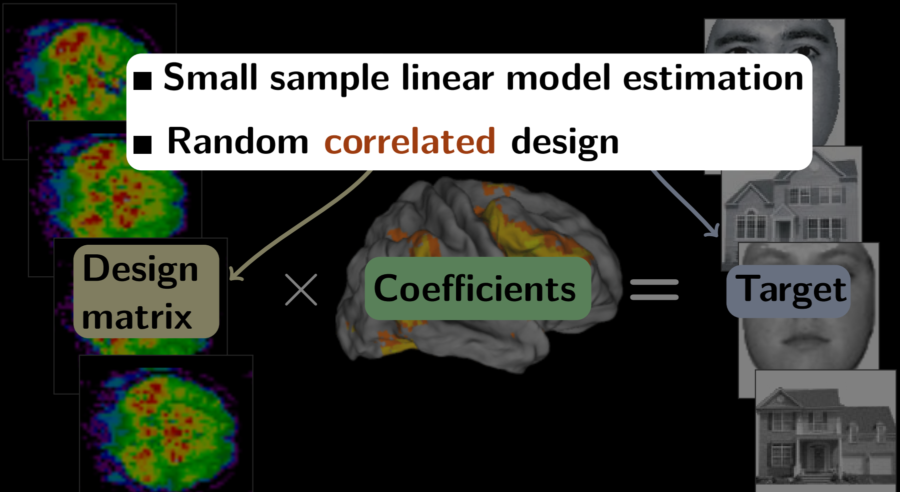
$=$

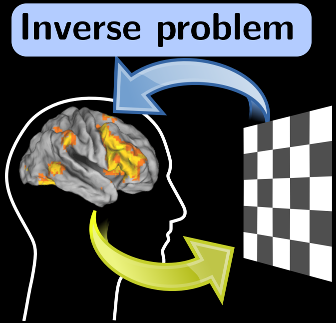
p $\sim$ 50 000

n $\sim$ 100 per category

4 Spatial penalties for learning from images

- Small sample linear model estimation
- Random correlated design





- Minimize an error term:

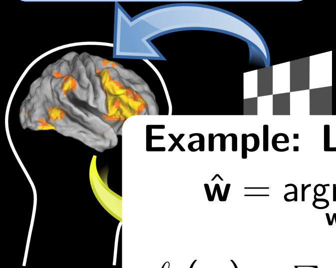
$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}}\, l(\mathbf{y} - \mathbf{X}\,\mathbf{w})$$

Ill-posed: $\mathbf{X}$ is not full rank

- Inject prior: regularize

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}}\, l(\mathbf{y} - \mathbf{X}\,\mathbf{w}) + p(\mathbf{w})$$

Inverse problem



- Minimize an error term:

$$\hat{\mathbf{w}} = \operatorname*{argmin}_{\mathbf{w}} l(\mathbf{y} - \mathbf{X}\,\mathbf{w})$$

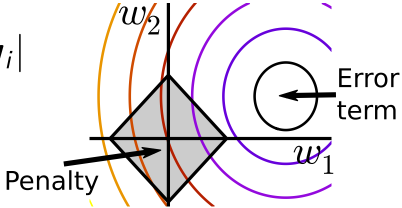
k

$+ \, p(\mathbf{w})$

Example: Lasso = sparse regression

$$\hat{\mathbf{w}} = \operatorname*{argmin}_{\mathbf{w}} \|\mathbf{y} - \mathbf{X}\,\mathbf{w}\|_2^2 + \ell_1(\mathbf{w})$$

$$\ell_1(\mathbf{w}) = \textstyle\sum_i |\mathbf{w}_i|$$



4 Good prediction $\neq$ good recovery

Simulations

Ground truth

Lasso

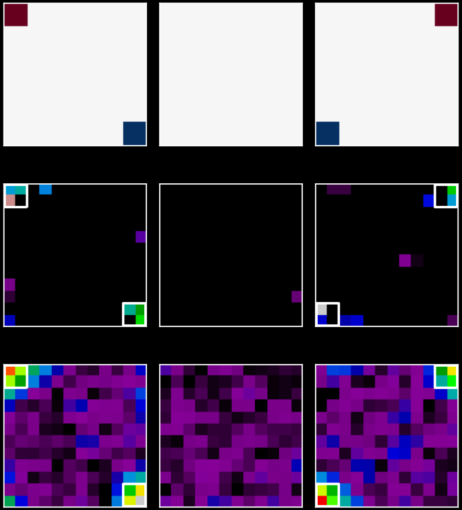
Prediction: 0.78

Recovery: 0.429

SVM

Prediction: 0.71

Recovery: 0.486



Need a method suited for recovery

4 Brain mapping & ℓ_1 sparse recovery

Recovering brain regions

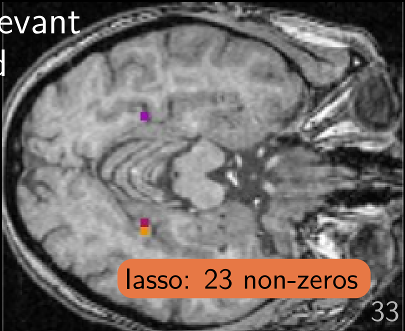
4 Brain mapping & ℓ_1 sparse recovery

Recovering k non-zero coefficients

- $n_{\text{min}} \sim 2\, k \log p$
- Restricted-isometry-like property:
  The design matrix is well-conditioned on sub-matrices of size $> k$
- Mutual incoherence:
  Relevant features S and irrelevant ones $\overline{S}$ are not too correlated

[Candes 2006]
[Tropp 2004]
[Wainwright 2009]

Violated by spatial correlations in our design



4 Randomized sparsity

[**Meinshausen and Buhlmann 2010**, Bach 2008]

- Perturb the design matrix:
  Subsample the data
  Randomly rescale features

  $+$ Run sparse estimator

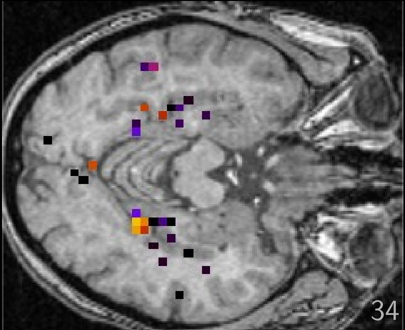
- Keep features that are often selected

$\Rightarrow$ **Good recovery without mutual incoherence**

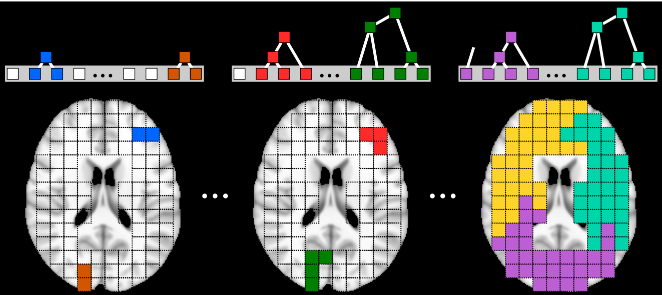
But RIP-like condition

Cannot recover large correlated groups ☹

For m correlated features,
selection frequency divided by m



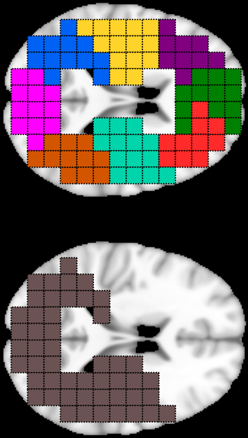
Spatially-connected hierarchical clustering
⇒ reduces voxel numbers [Michel Pat Rec 2011]



- Replace features by corresponding cluster average
  + Use a supervised learner on reduced problem

Cluster choice sub-optimal for regression

1st approach: randomized clustering



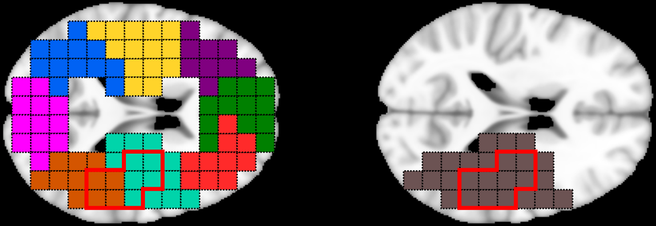
Combining

Clustering

Sparsity

[Varoquaux ICML 2012]

Hypothesis: clustering compatible with support($\mathbf{w}$)

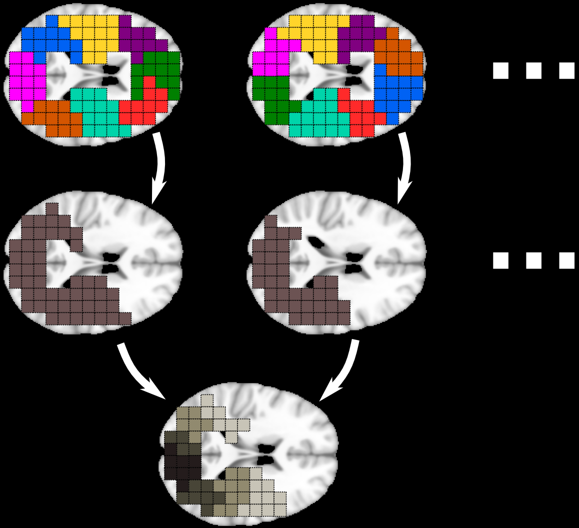


Benefits of clustering

- Reduced k and p
  $\Rightarrow$ $n > n_{\text{min}}$: good side of the "sharp threshold"
- Cluster together correlated features
  $\Rightarrow$ Improves RIP-like conditions

Recovery possible on reduced features

4 Randomized parcellations + sparsity



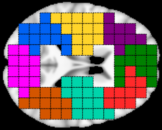
Randomization + Stability scores

- Marginalize the cluster choice
- Relaxes mutual incoherence requirement

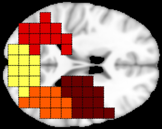
4 Algorithm

1 set `n_clusters` and sparsity by cross-validation

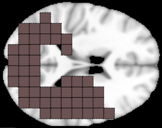
2 `loop:` perturb randomly data



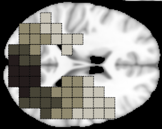
3 clustering to form reduced features



4 sparse linear model on reduced features



5 accumulate non-zero features



6 threshold map of apparition counts

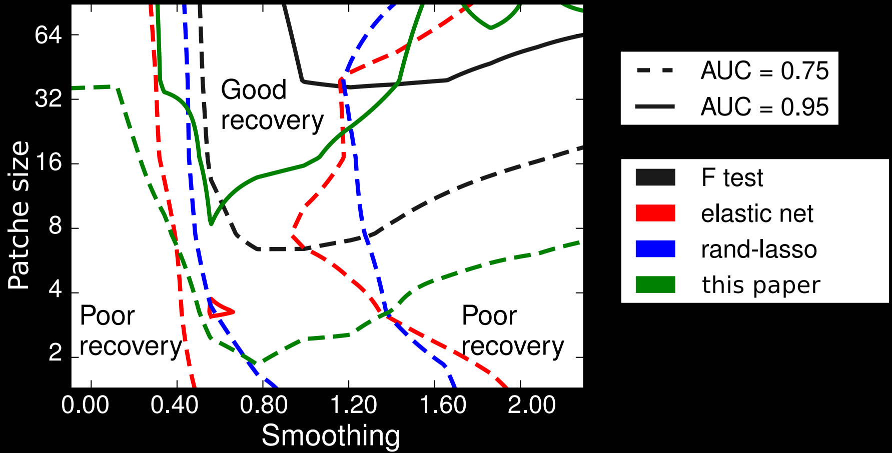
4 Simulations

- $p = 2048, k = 64, n = 256$ $(n_{\text{min}} > 1000)$
- Weights **w**: patches of varying size
- Design matrix: 2D Gaussian random images of varying smoothness

Estimators

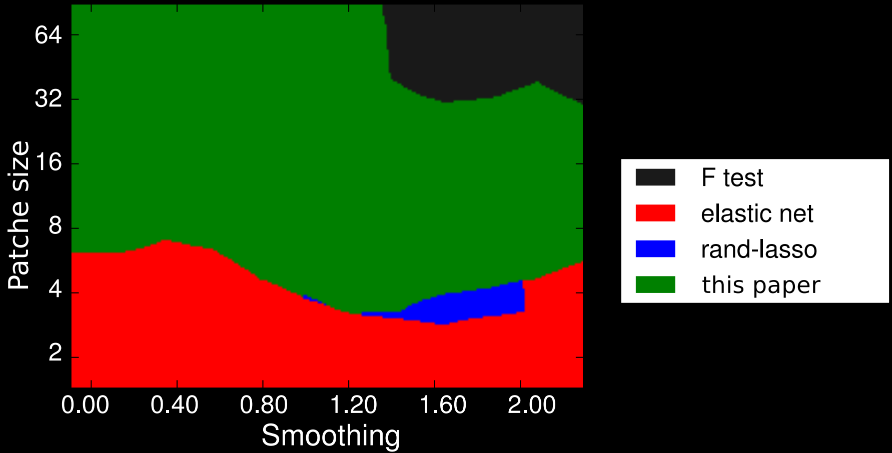
- Randomized lasso
- Elastic Net
- Our approach
- Univariate F test

4 When can we recover patches?



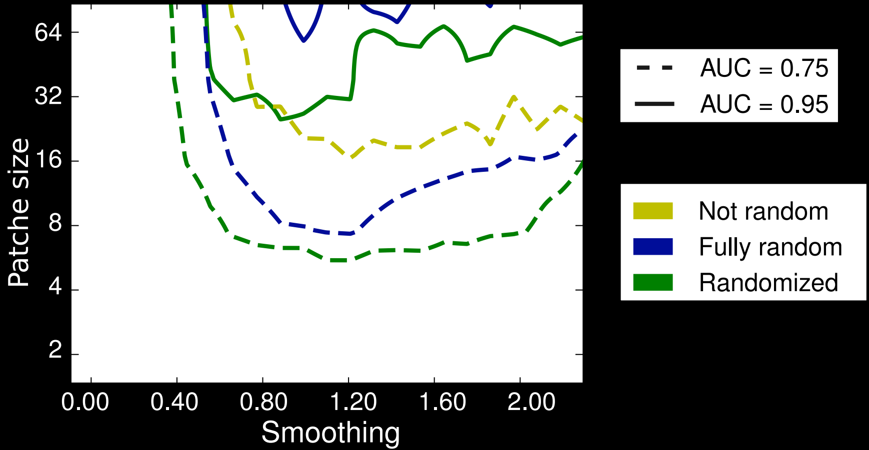
- Smoothness helps (reduces noise degrees of freedom)
- Small patches are hard to recover

4 What is the best method for patch recovery?



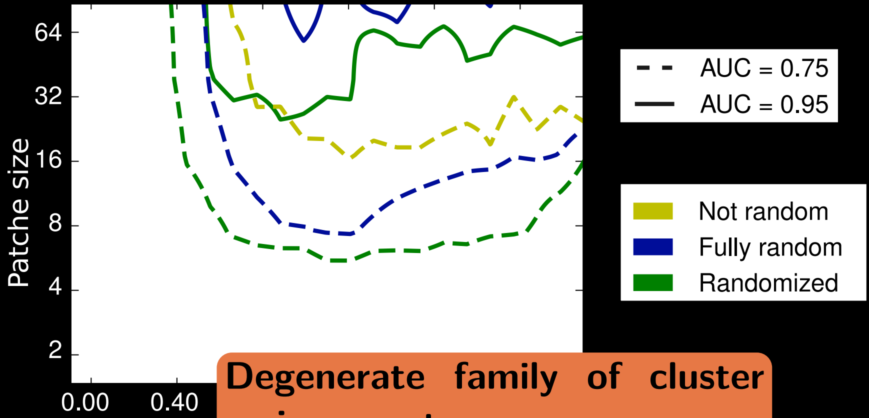
- For small patches: elastic net
- For large patches: randomized-clustered sparsity
- Large patches and very smooth images: F-test

4 Randomizing clusters matters!



- Non-random (Ward) clustering inefficient
- Fully-random performs quite well
- Randomized Ward gives an extra gain

4 Randomizing clusters matters!

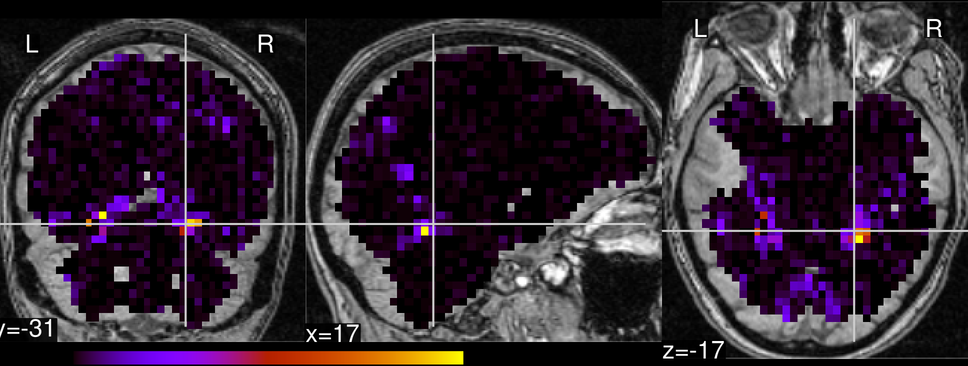


Degenerate family of cluster assignements

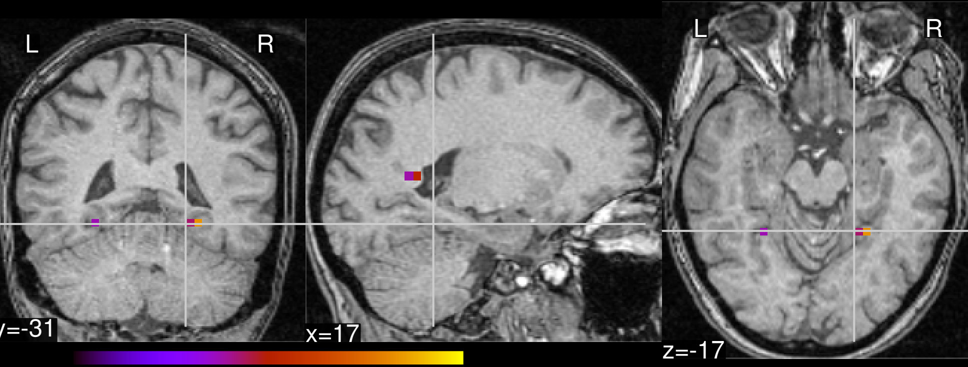
- Non-random (Ward) clustering inefficient
- Fully-random performs quite well
- Randomized Ward gives an extra gain

4 fMRI: face vs house discrimination [Haxby 2001]

Univariate F-scores

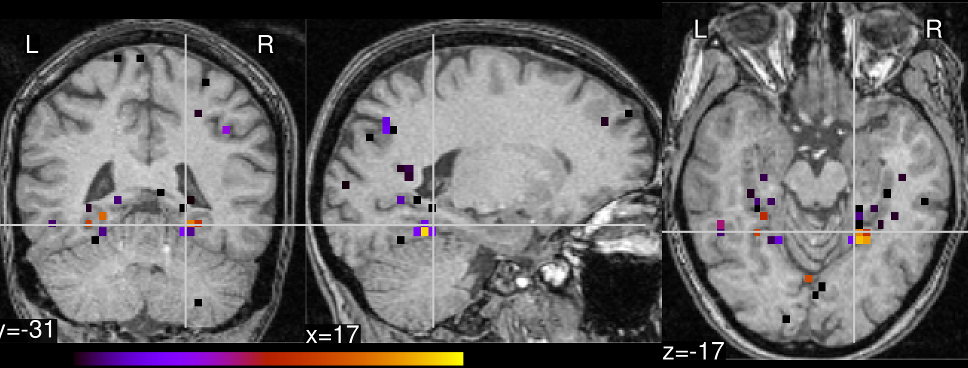


ℓ_1 Logistic



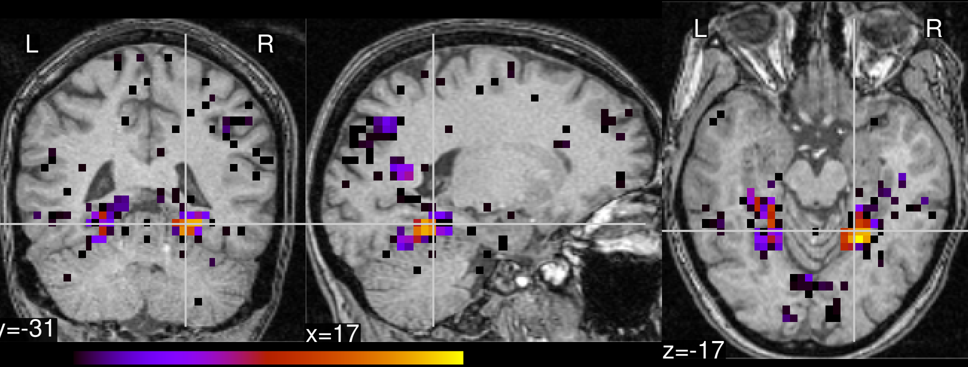
4 fMRI: face vs house discrimination [Haxby 2001]

Randomized ℓ_1 Logistic



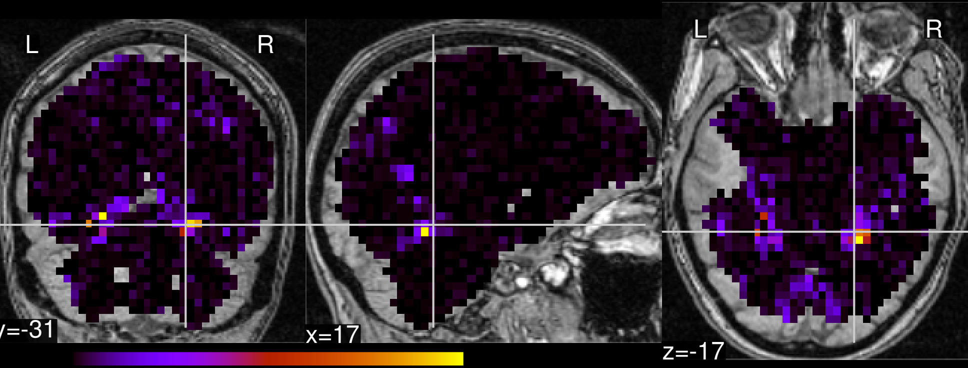
4 fMRI: face vs house discrimination [Haxby 2001]

Randomized Clustered ℓ_1 Logistic



4 fMRI: face vs house discrimination [Haxby 2001]

F-scores



1^{st} approach: randomized clustering

Sparse recovery of patches on spatially-correlated designs

Ingredients: **Clustering** + **Randomization**

⇒ Reduced feature set compatible with recovery:
matches sparsity pattern + recovery conditions

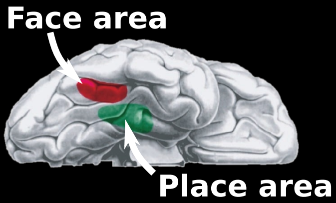
[Varoquaux ICML 2012]

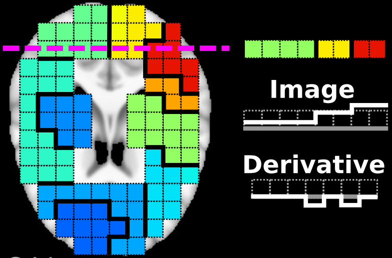
How to fit in compressive sensing theory?

2^{nd} approach: analysis sparsity

ℓ_1 penalty with an *analysis* operator: $p(\mathbf{w}) = \ell_1(\mathbf{K}\mathbf{w})$

- Spatial regularization necessary
- Neuroscientists think in terms of brain regions





Total-variation penalization

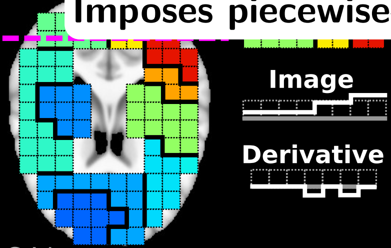
Impose sparsity on the gradient of the image:

$$p(\mathbf{w}) = \ell_1(\nabla \mathbf{w})$$

In fMRI: [Michel TMI 2011]

Original **Total-variation**

Imposes piecewise-smooth images



Impose sparsity on the gradient of the image:

$$p(\mathbf{w}) = \ell_1(\nabla \mathbf{w})$$

In fMRI: [Michel TMI 2011]

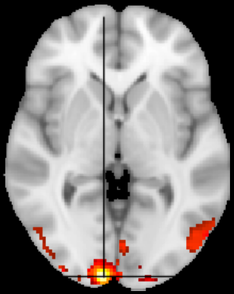
$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\text{argmin}}\ l(\mathbf{y} - \mathbf{X}\,\mathbf{w}) + TV(\mathbf{w})$$

- l: least-square or logistic-regression
- p: TV: isotropic total variation: $\ell_{21}(\nabla w)$

Prediction performance:

Feature screening + SVC	0.77
Sparse regression	0.78
Total Variation	**0.84**

(explained variance)

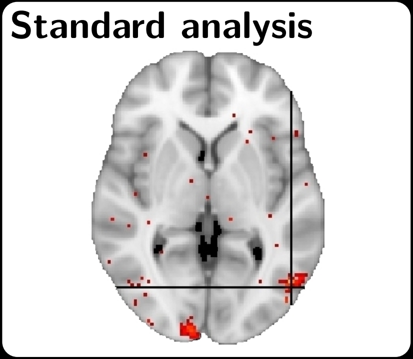


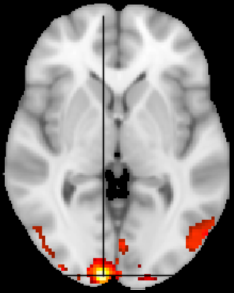
[Michel TMI 2011]

4 fMRI decoding with TV

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}}\, l(\mathbf{y} - \mathbf{X}\,\mathbf{w}) + TV(\mathbf{w})$$

- l: least-square or logistic-regression
- p: TV: isotropic total variation: $\ell_{21}(\nabla w)$





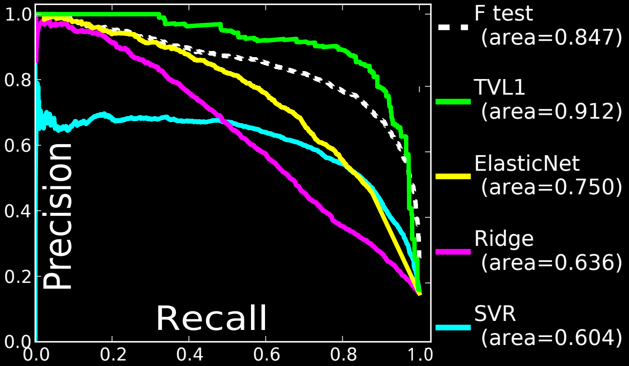
[Michel TMI 2011]

4 fMRI decoding with TV-ℓ_1

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}}\, l(\mathbf{y} - \mathbf{X}\,\mathbf{w}) + TV(\mathbf{w}) + \ell_1(w)$$

- Adding ℓ_1 = extending analysis operator
- Retreives sparsity in the original coefficients

On simulations: sparse recovery precision-recall

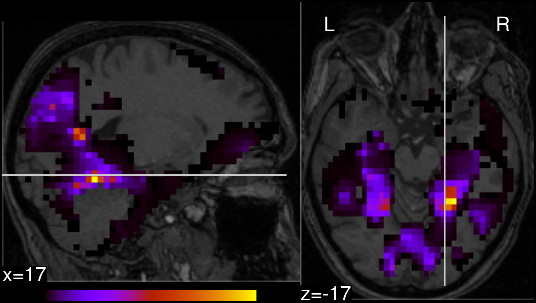


[Gramfort 2013]

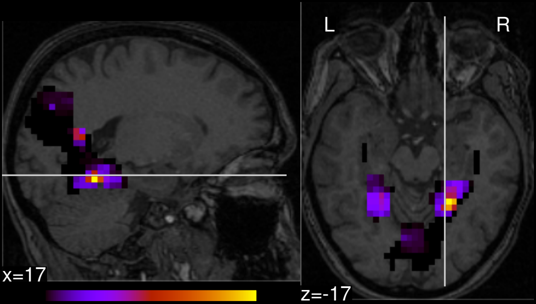
4 fMRI decoding with TV-ℓ_1: implementation

Convergence matters

Stopping: $\Delta E < 10^{-1}$

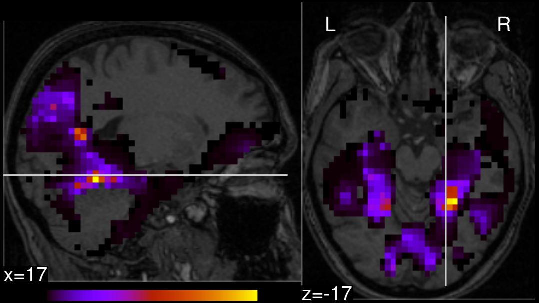


Stopping: $\Delta E < 10^{-3}$

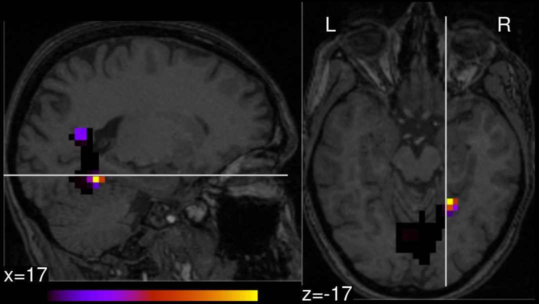


Convergence matters

Stopping: $\Delta E < 10^{-1}$



Stopping: $\Delta E < 10^{-5}$



Optimization algorithms: FISTA

- FISTA loop for regression:

  1. Gradient descent on the datafit term

  2. Proximal operator for TV,
     computed with an inner FISTA loop

Bottleneck: Gradient descent step costly

$$\hat{\mathbf{w}} = \operatorname*{argmin}_{\mathbf{w}} \|\mathbf{y} - \mathbf{X}\,\mathbf{w}) + TV(\mathbf{w}\|_2^2 + \ell_1(w)$$

Gradient: $\mathbf{X}^t\,\mathbf{X}\,\mathbf{w}$ with $\mathbf{X}$ big and dense

[Domathob PRNI 2014]

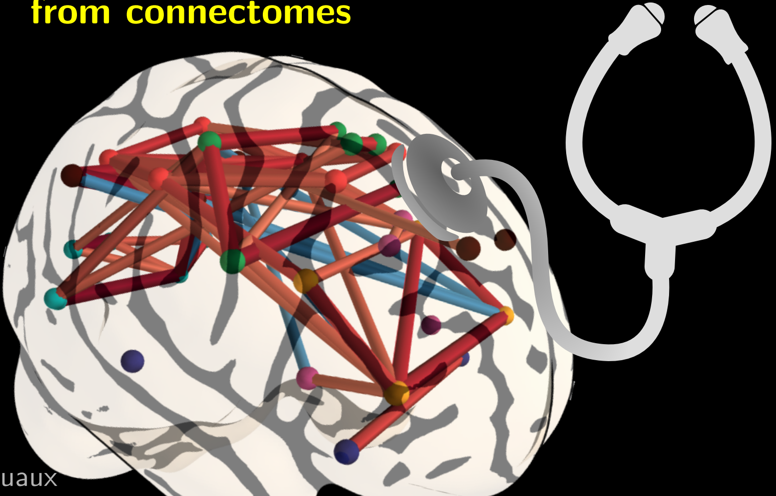
Take home messages

- Recovery $\neq$ prediction
- Univariate tests work very well to recover

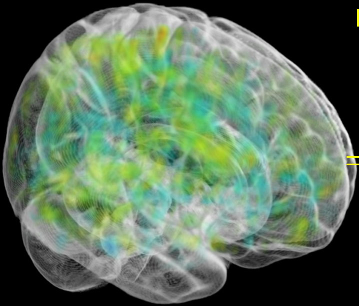
  Also in genomics [Haury PLOS One 2011]
- Some form of spatial regularization usefull
- Speed matters! $\mathbf{X}: (100\,000, 600)$

5 Learning from spontaneous activity

from connectomes



5 Resting-state: Spontaneous activity

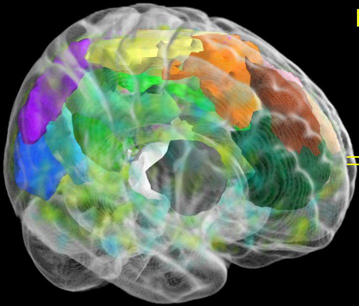


- In the absence of explicit tasks cognitive circuits are recruited spontaneously

$\Rightarrow$ Meaningful cofluctuation patterns

Diagnostic interest for disabled patients

5 Resting-state: Spontaneous activity



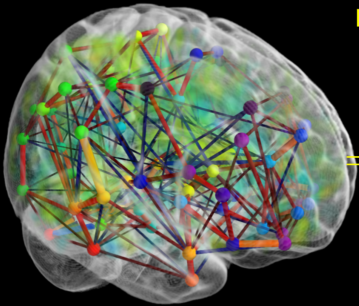
- In the absence of explicit tasks cognitive circuits are recruited spontaneously

$\Rightarrow$ Meaningful cofluctuation patterns

Diagnostic interest for disabled patients

1. Learn spatial maps/regions – *Unsupervised learning*

5 Resting-state: Spontaneous activity

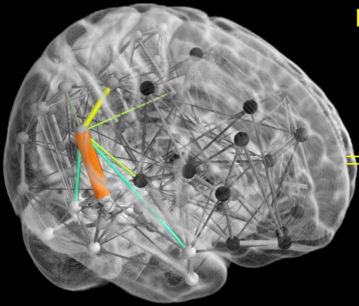


- In the absence of explicit tasks cognitive circuits are recruited spontaneously

$\Rightarrow$ Meaningful cofluctuation patterns

Diagnostic interest for disabled patients

1. Learn spatial maps/regions – *Unsupervised learning*
2. Learn interaction graph – *Unsupervised learning*



- In the absence of explicit tasks cognitive circuits are recruited spontaneously

$\Rightarrow$ Meaningful cofluctuation patterns

Diagnostic interest for disabled patients

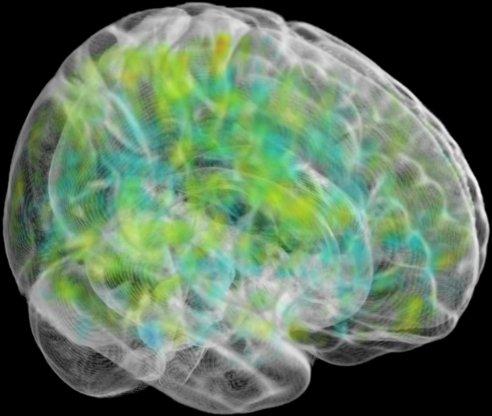
1. Learn spatial maps/regions – *Unsupervised learning*
2. Learn interaction graph – *Unsupervised learning*
3. Inter-subject/inter-condition prediction – *Supervised learning*

[Varoquaux MICCAI 2010, NIPS 2010, 2011 IPMI, ...]

Working hypothesis:

Observing linear mixtures of networks at rest

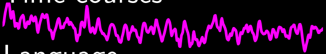
Time courses



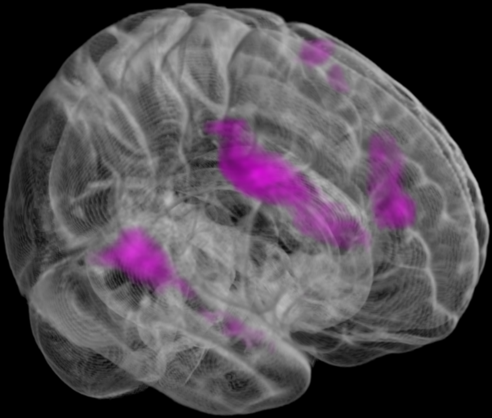
Working hypothesis:

Observing linear mixtures of networks at rest

Time courses



Language

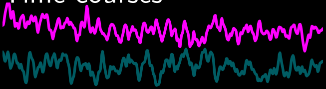


5 Mixture models: linear decompositions

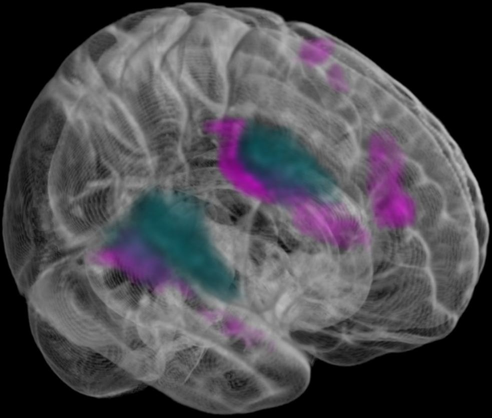
Working hypothesis:

Observing linear mixtures of networks at rest

Time courses

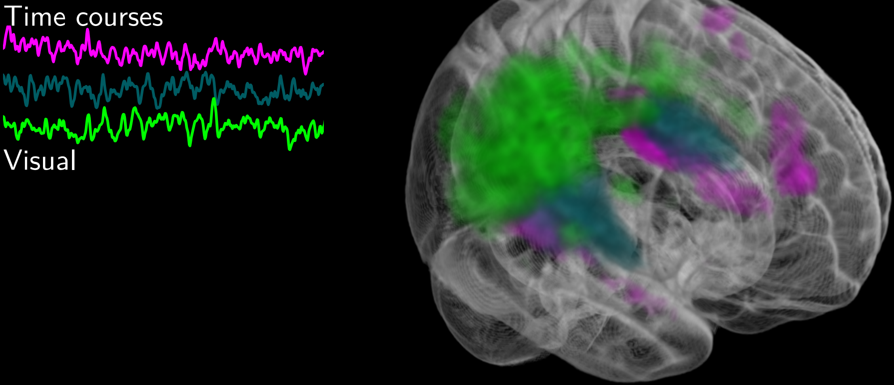


Audio



Working hypothesis:

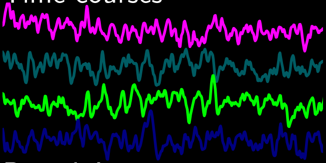
Observing linear mixtures of networks at rest



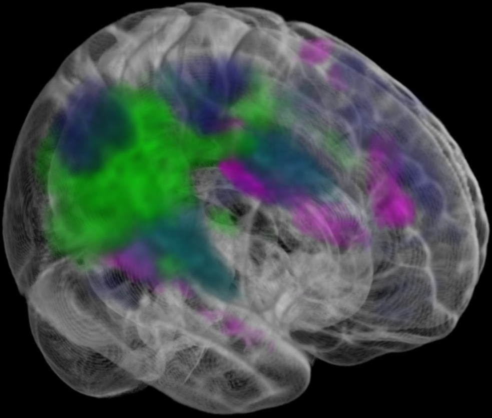
Working hypothesis:

Observing linear mixtures of networks at rest

Time courses

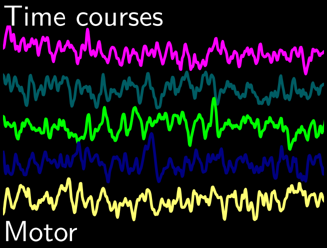


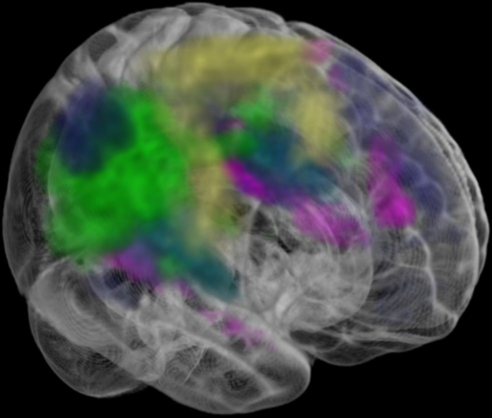
Dorsal Att.



Working hypothesis:

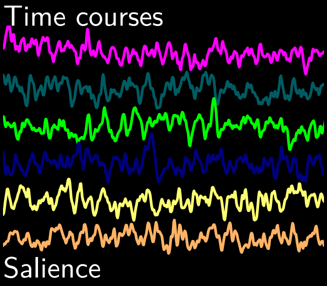
Observing linear mixtures of networks at rest

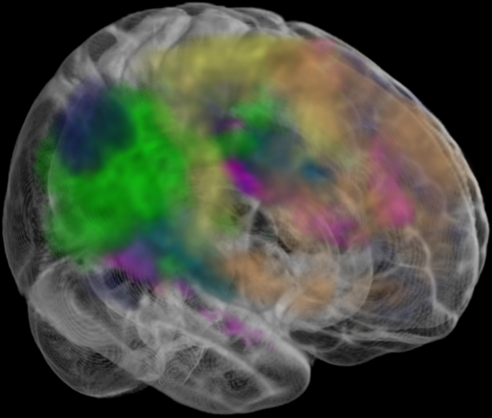




Working hypothesis:

Observing linear mixtures of networks at rest

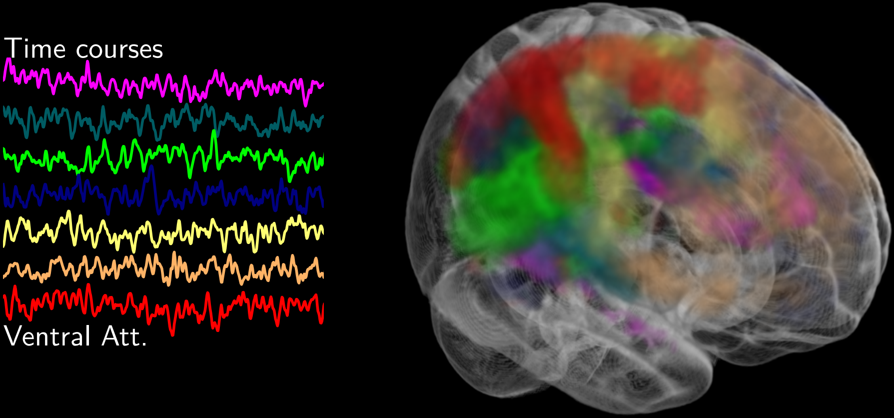




5 Mixture models: linear decompositions

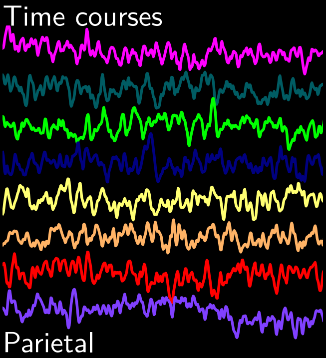
Working hypothesis:

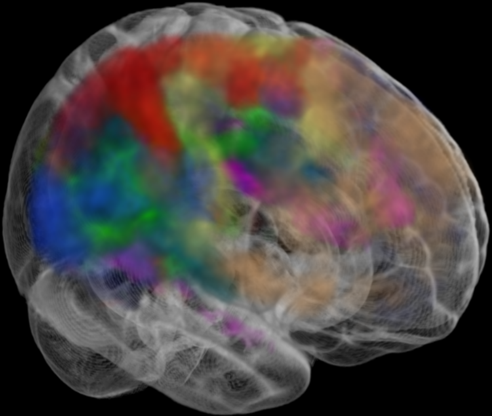
Observing linear mixtures of networks at rest



Working hypothesis:

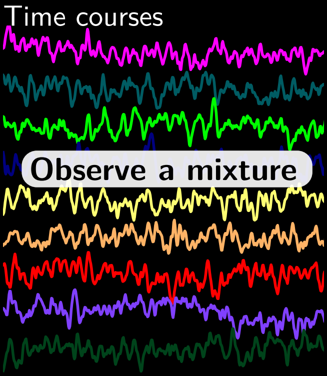
Observing linear mixtures of networks at rest

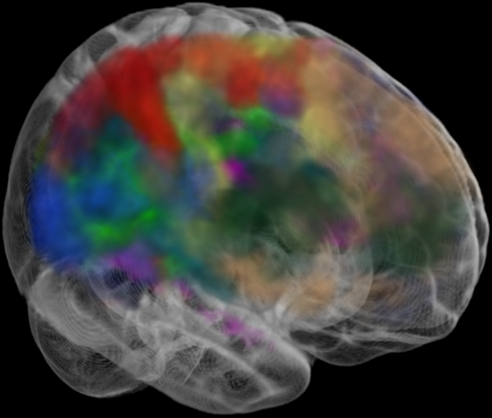




Working hypothesis:

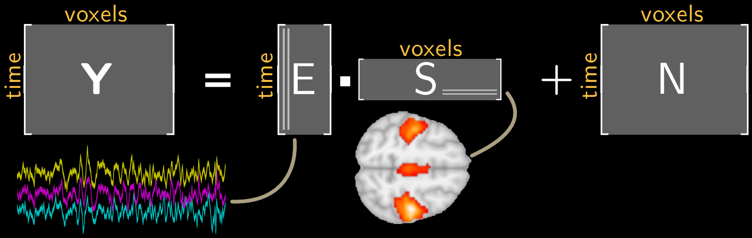
Observing linear mixtures of networks at rest





How to unmix networks?

Segmenting regions from spontaneous activity



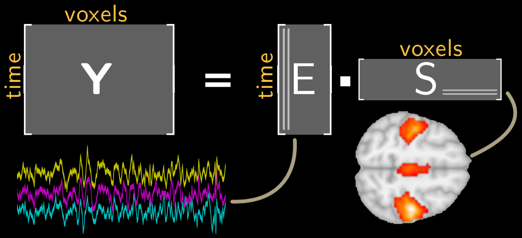
Linear decomposition model

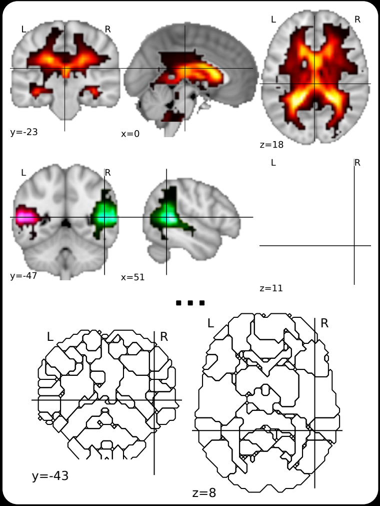
- spatial maps, **S**
- residuals, **N**
- time series, **E**

Independant Component Analysis

[Beckman TMI 2004, Varoquaux NeuroImage 2010]

5 Segmenting regions from spontaneous activity





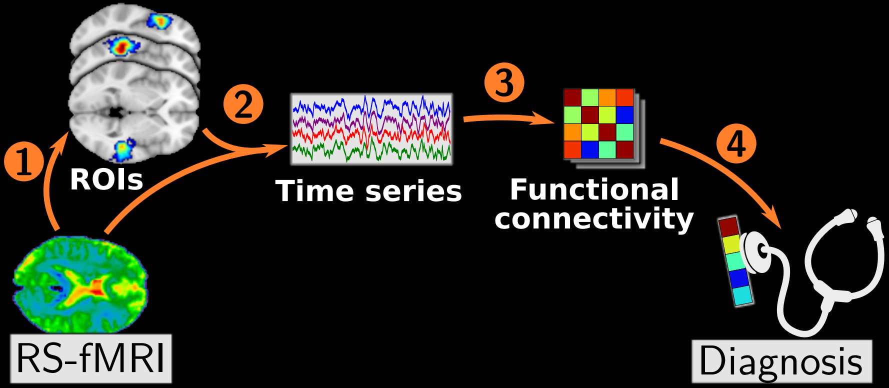
Linear decomposition model

- spatial maps, **S**
- residuals, **N**
- time series, **E**

TV-ℓ_1 penalty on maps S

[Abraham MICCAI 2013]

5 The connectome classification pipeline



Preliminary results on Autism
Leave-one-site out cross-validation
$\Rightarrow$ 72% accuracy

6 Beyond equations: software

- How to we reach our target audience (neuroscientists)?

- How do we disseminate our ideas?

- How do we facilitate new ideas?

6 Python as a scientific environment

- General purpose
- Easy, readable syntax
- Interactive (ipython)
- Great scientific libraries (numpy, scipy, matplotlib...)

- Code lines are costly

$\Rightarrow$ Open source + community driven

Need quality and impact

$\Rightarrow$ Focus on the general purpose libraries first

Scikit-learn: machine learning in Python

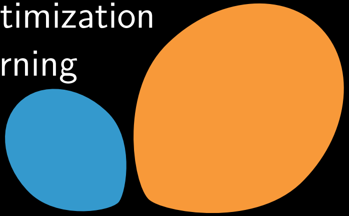
`http://scikit-learn.org`

[Pedregosa 2011]

Computational performance

	scikit-learn	mlpy	pybrain	pymvpa	mdp	shogun
SVM	**5.2**	9.47	17.5	11.52	40.48	5.63
LARS	**1.17**	105.3	-	37.35	-	-
Elastic Net	**0.52**	73.7	-	1.44	-	-
kNN	0.57	1.41	-	**0.56**	0.58	1.36
PCA	**0.18**	-	-	8.93	0.47	0.33
k-Means	1.34	0.79	∞	-	35.75	**0.68**

- Algorithms rather than low-level optimization
  convex optimization + machine learning
- Avoid memory copies



Community

- 200 contributors since 2008, 1500 github forks
- 25 contributors in latest release (3 months span)

Why this success?

- Trendy topic?
- Low barrier of entry
- Friendly and very skilled mailing list
- Credit to people

Factor 10 in time investment

- Corner cases in algorithm (numerical stability)
- Multiple platforms and library versions (Blas ☹)
- Documentation
- Making it simpler (and get less educated users)
- User and developer support ($\sim$ 100 mails/day)

Exhausting,
but has impact on science and society

Factor 10 in time investment

Technical + scientific tradeoffs

- Ease of install/ease of use rather than speed
- Focus on “old science”
- Nice publications and theorems are not a recipe for useful code

Exhausting,
but has impact on science and society

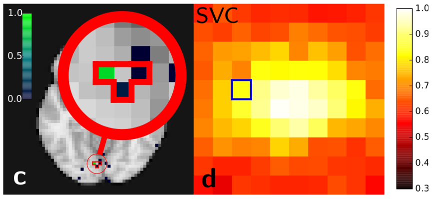
Project scope

Very preliminar

Machine learning for neuroimaging:
make using scikit-learn on neuroimaging easy

The target user base is small ☹

Examples in the docs

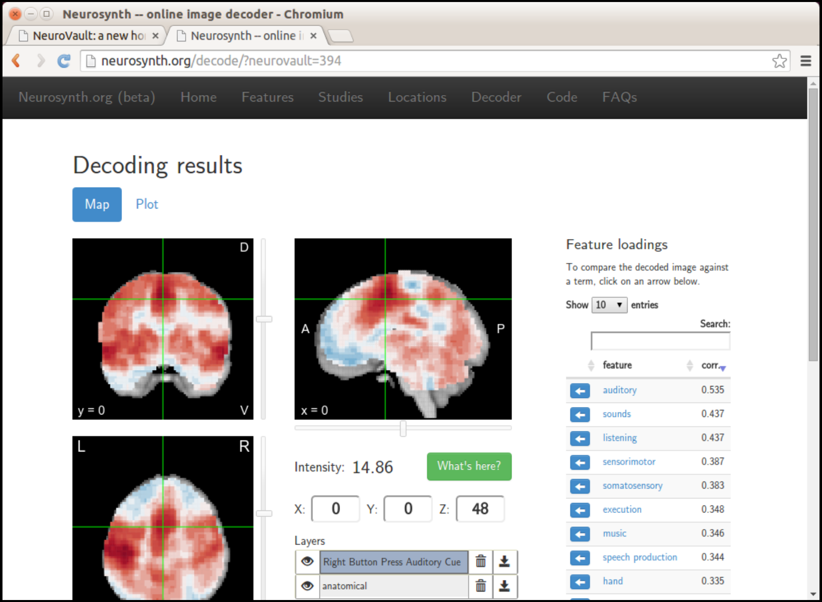


Data from Miyawaki 2008

- Run out of the box, downloading **open data**
- Produce a clear figure

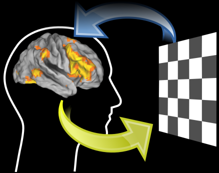
Routine, simple, reproduction of papers

6 NeuroSynth + Neurovault: decoding as a service



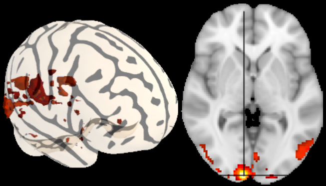
Machine learning for brain imaging

Statistical learning to study brain function



- Learning problems, but not only about prediction error

- Spatial regularization for linear models
  Total variation + randomization



- Validation is very hard
  All model are wrongs, and data is scarce

- Speed and parameter selection matter
  Users will not adapt

Positions available

Bibliography

- [Abraham 2013] A. Abraham *el al.*, *Extracting brain regions from rest fMRI with Total-Variation constrained dictionary learning*, Med Imag Comp Aided Intervention (2013)
  `http://hal.inria.fr/hal-00839984/en`
- [Beckmann 2004] *Probabilistic Independent Component Analysis for functional Magentic Resonance Imaging*, IEEE Transactions in medical imaging 2004
- [Dohmatob 2014] E. Dohmatob *el al.*, *Benchmarking solvers for TV-l1 least-squares and logistic regression in brain imaging*, Pattern Recognition in Neuro Imaging (2014)
  `http://hal.inria.fr/hal-00991743/en`
- [Gramfort 2013] A. Gramfort, *et al.*, *Identifying predictive regions from fMRI with TV-L1 prior*, Pattern Recognition in Neuro Imaging (2013)
  `http://hal.inria.fr/hal-00839984/en`

References (not exhaustive)

- [Haury 2011] A-C. Haury, *el al.*, *The Influence of Feature Selection Methods on Accuracy, Stability and Interpretability of Molecular Signatures*, PLoS ONE, (2011), e28210

- [Michel TMI 2011] V. Michel, *et al.*, *Total variation regularization for fMRI-based prediction of behaviour*, IEEE Transactions in medical imaging (2011)
`http://hal.inria.fr/inria-00563468/en`

- [Michel Pat Rec 2011] V. Michel, *et al.*, *A supervised clustering approach for fMRI-based inference of brain states*, Pattern Recognition (2011)
`http://hal.inria.fr/inria-00589201/en`

- [Smith 2010] *Network Modelling Methods for fMRI*, NeuroImage

- [Pedregosa ICML 2011] F. Pedregosa, *el al.*, *Scikit-learn: machine learning in Python*, JMRL (2011)
`http://hal.inria.fr/hal-00650905/en`

References (not exhaustive)

- [Poldrack 2011] R.A. Poldrack, *el al.*, *Handbook of functional MRI data analysis*, Cambridge University Press (2011)
- [Varoquaux MICCAI 2010] *Detection of brain functional-connectivity difference in post-stroke patients using group-level covariance modeling*, Med Imag Comp Aided Intervention http://hal.inria.fr/inria-00512417/en
- [Varoquaux NIPS 2010] *Brain covariance selection: better individual functional connectivity models using population prior*, Neural Inf Proc Sys http://hal.inria.fr/inria-00512451/en
- [Varoquaux IPMI 2011] *Multi-subject dictionary learning to segment an atlas of brain spontaneous activity*, IPMI http://hal.inria.fr/inria-00588898/en

References (not exhaustive)

- [Varoquaux ICML 2012] G. Varoquaux, *et al*, *Small-sample brain mapping: sparse recovery on spatially correlated designs with randomization and clustering*, ICML (2012)
  http://hal.inria.fr/hal-00705192/en

- [Varoquaux 2013] *Learning and comparing functional connectomes across subjects*, NeuroImage
  http://hal.inria.fr/hal-00812911/en