Challenges and limitations of classifiers for analyzing fMRI data

Francisco Pereira Computer Science Department
Center for the Neural Basis of Cognition
Tom Mitchell Machine Learning Department
Carnegie Mellon University

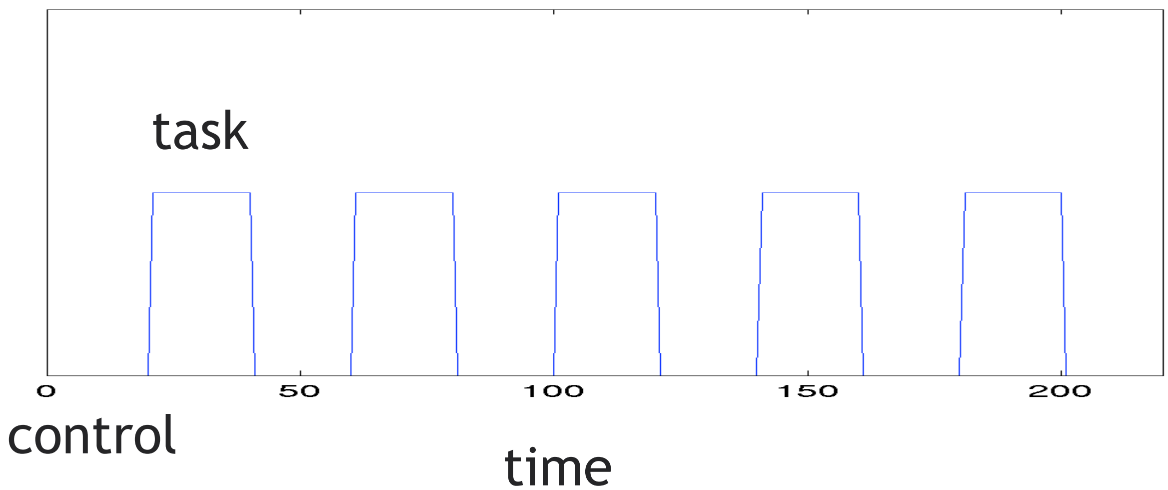
[fMRI data from Marcel Just and collaborators,
Center for Cognitive Brain Imaging, CMU]

fMRI analysis

A typical experiment is designed to have the subject perform:

- a task of interest (e.g. read a word)
- a control task (e.g. read a nonsense word)

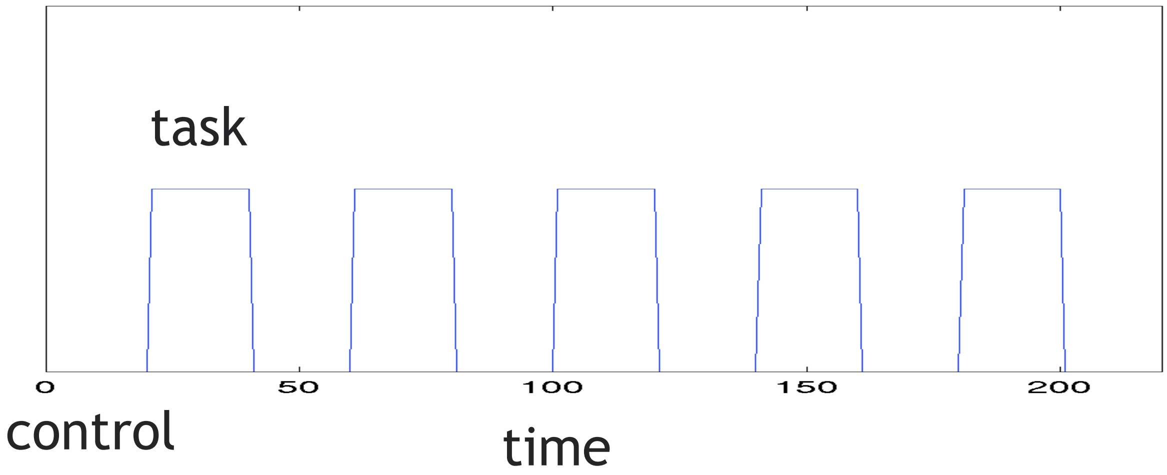
experimental conditions



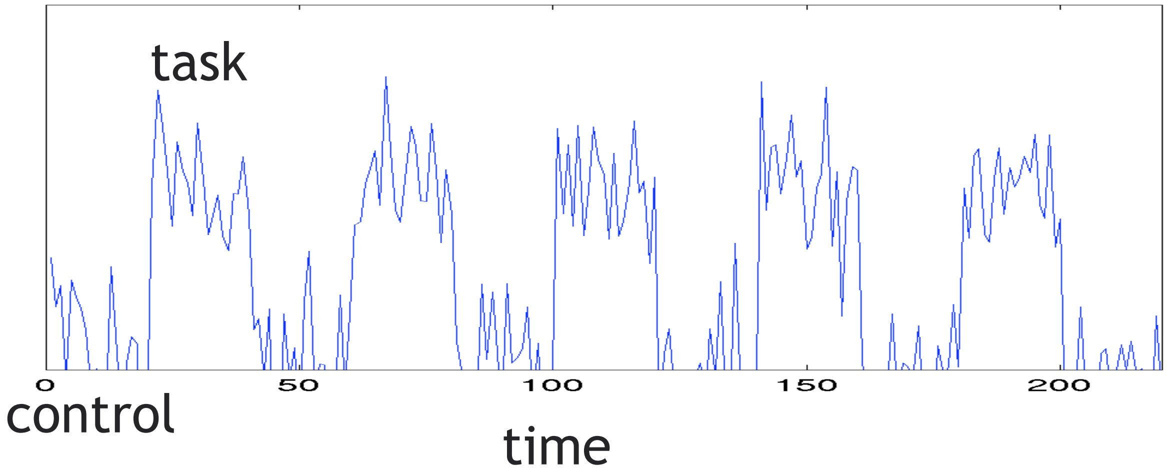
reference
time series

fMRI analysis

The goal is to find voxels that match the reference



reference
time series

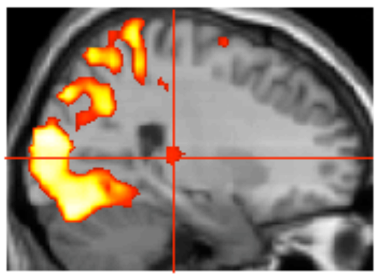


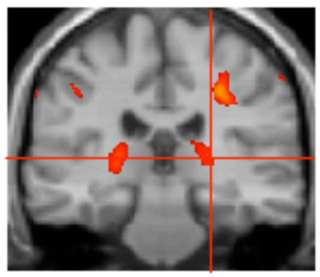
voxel
time series

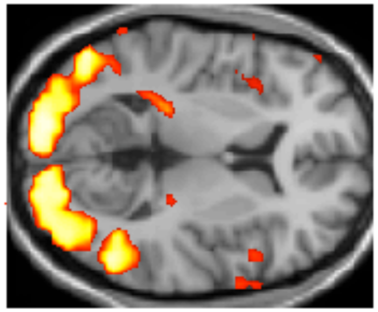
fMRI analysis

This is done for each voxel in the brain

- yields an image with the matching score for each voxel
- that image is thresholded leaving only significant matches





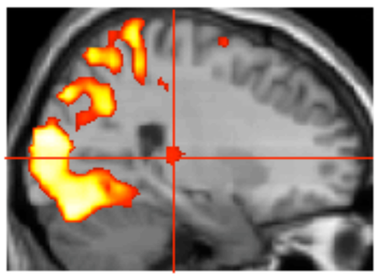


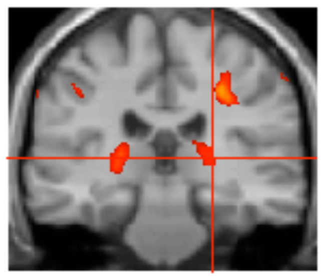
statistical parametric map (SPM)

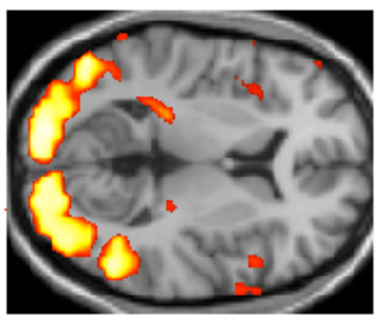
fMRI analysis

This is done for each voxel in the brain

- yields an image with the matching score for each voxel
- that image is thresholded leaving only significant matches







statistical parametric map (SPM)

a.k.a. BRAIN BLOBS

fMRI analysis

SPM as an instrument

- identifies voxels more active in task than in control
- tests statistical significance of what was identified
- location

  "which voxels are more active in task than in control images?"

fMRI analysis

SPM as an instrument

- identifies voxels more active in task than in control
- tests statistical significance of what was identified
- location

  "which voxels are more active in task than in control images?"
- location

  "is the location of active voxels reliable across subjects?"

fMRI analysis

SPM as an instrument

- identifies voxels more active in task than in control
- tests statistical significance of what was identified
- location

  "which voxels are more active in task than in control images?"
- location

  "is the location of active voxels reliable across subjects?"
- location

  "does the location make sense in the light of prior knowledge?"

fMRI analysis

- if you can only test for location, experimental hypotheses will be formulated in terms of location
- ever finer contrasts...

fMRI analysis

- if you can only test for location, experimental hypotheses will be formulated in terms of location
- ever finer contrasts...

"Brain Activation During Viewing of Erotic Film Excerpts under Influence of Alcohol"

"In order to examine this issue, functional MRI was performed in a group of young, healthy, right handed males. Subjects viewed erotic film excerpts alternating with emotionally neutral excerpts in a standard block-design paradigm."

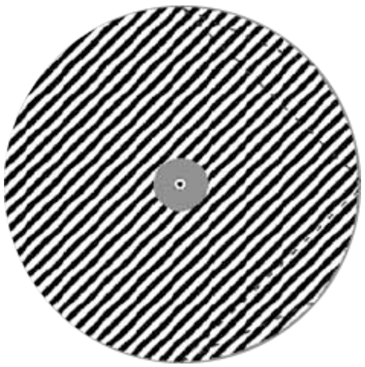
fMRI analysis

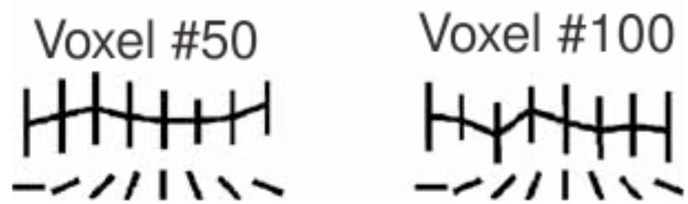
What could be missing?

- voxel interactions
- very small/unreliable differences between conditions
- making sense of many task conditions
- ...

fMRI analysis with classifiers

[Kamitani&Tong, 2005]





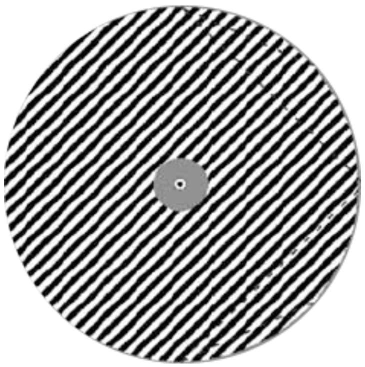
voxel responses
orientations

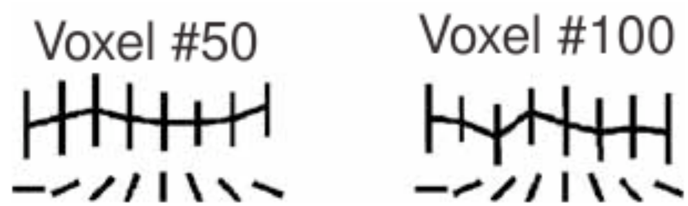
subjects see gratings in one of 8 orientations

voxels in visual cortex respond similarly to different orientations

fMRI analysis with classifiers

[Kamitani&Tong, 2005]



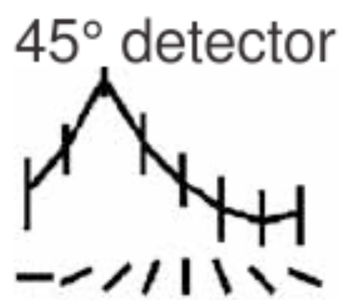


voxel responses
orientations

subjects see gratings in one of 8 orientations

voxels in visual cortex respond similarly to different orientations

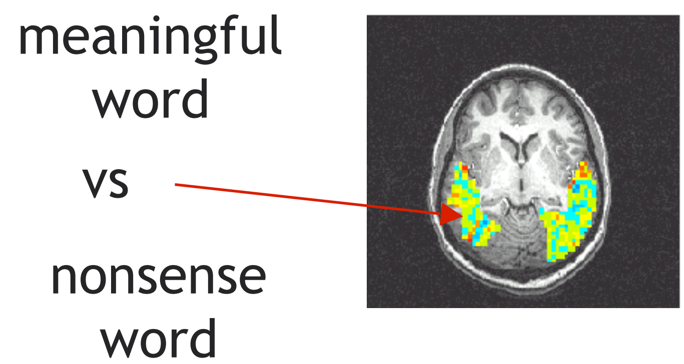
yet, voxels can be combined to predict the orientation of the grating being seen!



what questions can we ask?

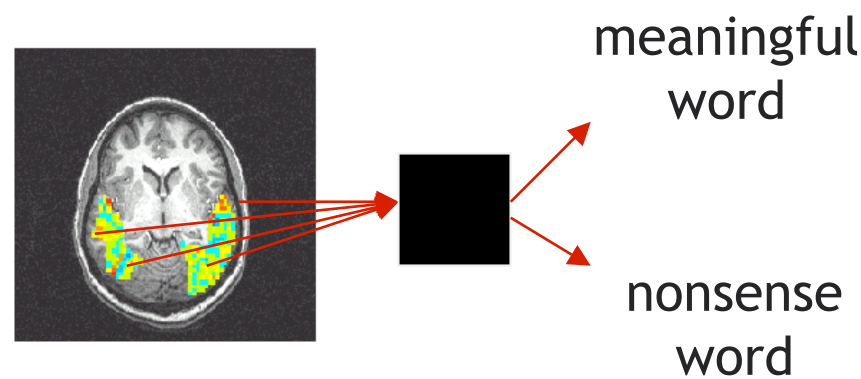
univariate:

Is the activity of voxel *v* sensitive to an experimental condition?



multivariate:

Can voxel set $S=\{v_1, \ldots v_n\}$ be used to predict the experimental condition?



what questions should we ask?

- Can we predict?
- Can we say what in the image is related to what we are trying to predict, and how?
- Can we use prior knowledge to make better classifiers?
- Can we test hypotheses?

Exploratory

Confirmatory

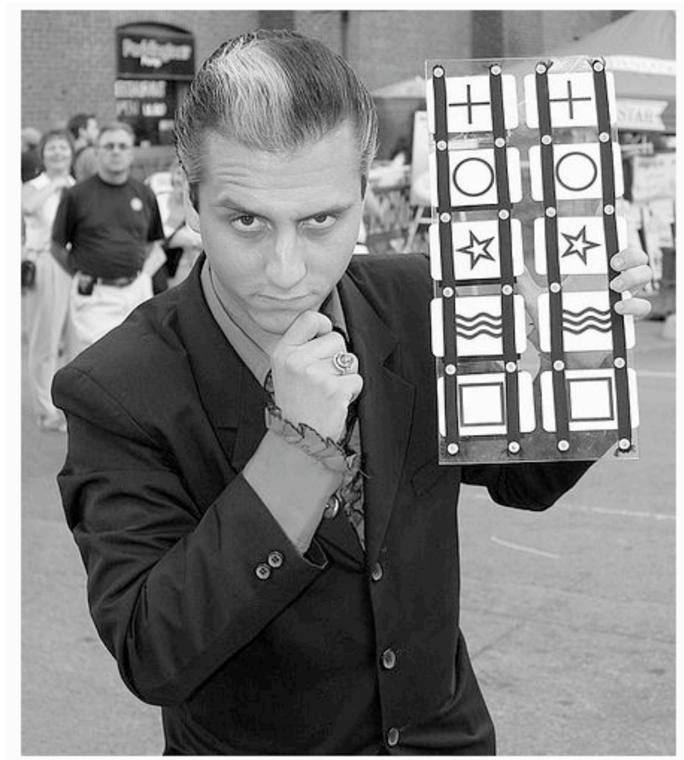
can we predict?

[Mitchell et al 2004, Haynes 2006, Norman 2006]

- is the subject seeing a sentence or a picture?
- which of several categories of words or pictures is a subject seeing?
- is the subject reading an ambiguous sentence?
- will the subject answer correctly?
- what is the orientation of a stimulus visual grating?
- is there a face/music/tools/... in a film clip being seen?
- what is the subject perceiving?
- is the subject concealing information?

yes, one can read minds*...

*Conditions may apply



… but it comes at a price

Why?

- Few examples (10s-100s)
- Many features (10K-100K)

… but it comes at a price

Why?

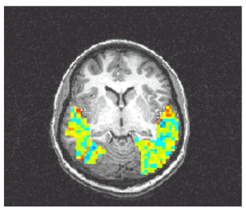
- Few examples (10s-100s)
- Many features (10K-100K)
- Noise:
  - the scanner
  - the body/brain
  - the subject
  - the subject
  - the subject

  - from our viewpoint: spatially correlated, heavy-tailed

so what?

Common to almost all papers:

- Features are voxels



$\underbrace{\boxed{}\boxed{}\boxed{}\boxed{}\boxed{}\boxed{}}_{\text{voxels (features)}}$

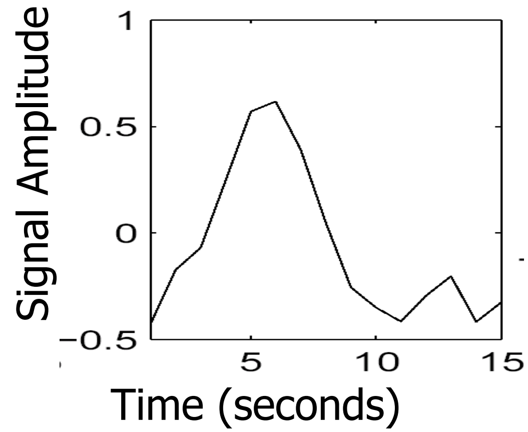
- Linear discriminant classifiers

If weight 0 + weight1 x voxel 1 + weight2 x voxel 2 + … + + weight n x voxel n > 0 tools
otherwise buildings

so what?

Common to almost all papers:

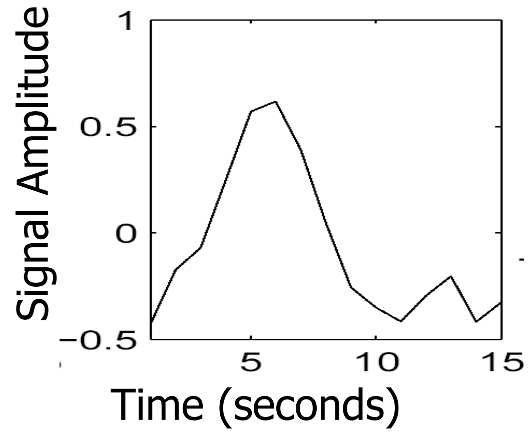
- Examples are not individual images
  - response to short neural activity is long
  - responses add up
  - easier to average over time



so what?

Common to almost all papers:

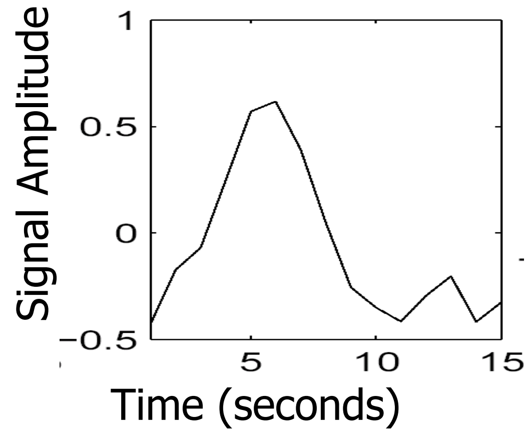
- Examples are not individual images
  - response to short neural activity is long
  - responses add up
  - easier to average over time
- Need for voxel selection
  - activation profile
  - accuracy/mutual information with target variable
  - location



so what?

Common to almost all papers:

- Examples are not individual images
  - response to short neural activity is long
  - responses add up
  - easier to average over time
- Need for voxel selection
  - activation profile
  - accuracy/mutual information with target variable
  - location



- If a classifier can predict, the selection criterion identifies voxels related to the target ...
- ... but what does the classifier itself tell us?

experiments

- Studies designed to:
  - elicit mental representations of semantic categories
  - try to understand how those map to brain activation

experiments

- Studies designed to:
  - elicit mental representations of semantic categories
  - try to understand how those map to brain activation

- The features are voxels
- Linear discriminant classifiers
- Cross-validation
- Best subject results (consistent across subjects)

2 categories experiment

- Subjects read concrete nouns in 2 categories
  - words are either tools or buildings
  - task:

    see a word/think about it for 3 sec., 8 sec. pause afterwards
  - e.g. “hammer”, “saw”, “palace”, “hut”

2 categories experiment

- Subjects read concrete nouns in 2 categories
  - words are either tools or buildings
  - task:

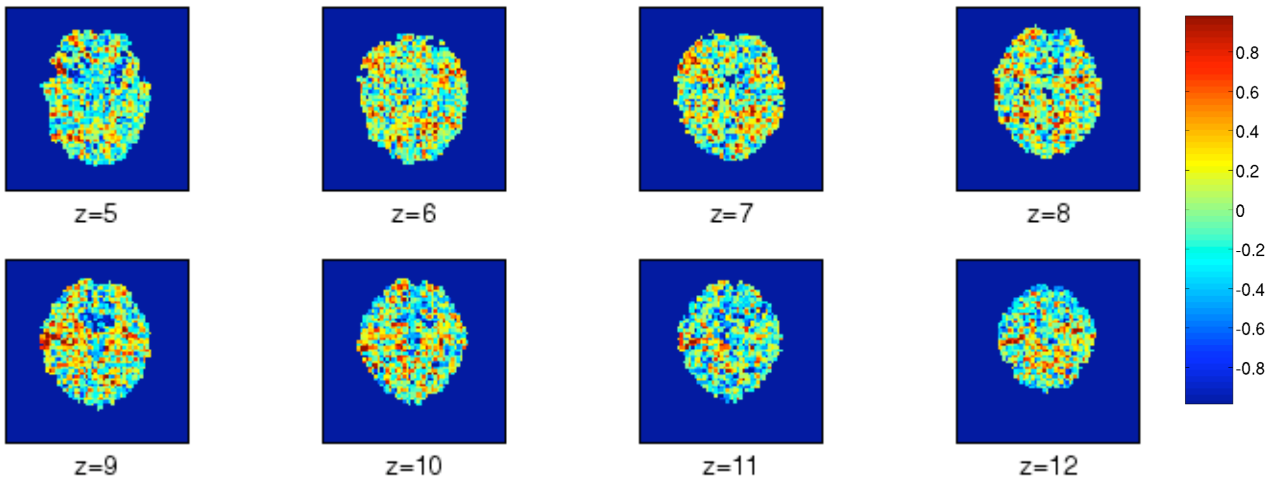
    see a word/think about it for 3 sec., 8 sec. pause afterwards
  - e.g. “hammer”, “saw”, “palace”, “hut”

- Classification task: predict the category
- Example:

  average 3D image of middle 4 secs of a trial
- 42 examples of each noun category
- 10K-20K features

2 categories linear discriminants

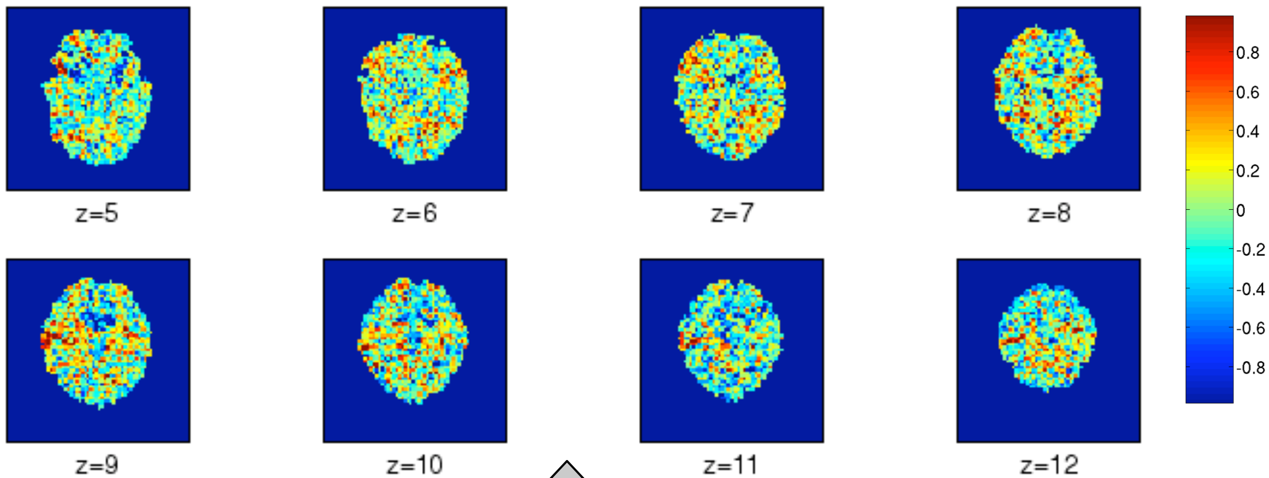
It's possible to predict category using all the voxels



GNB weights
(accuracy 65%)

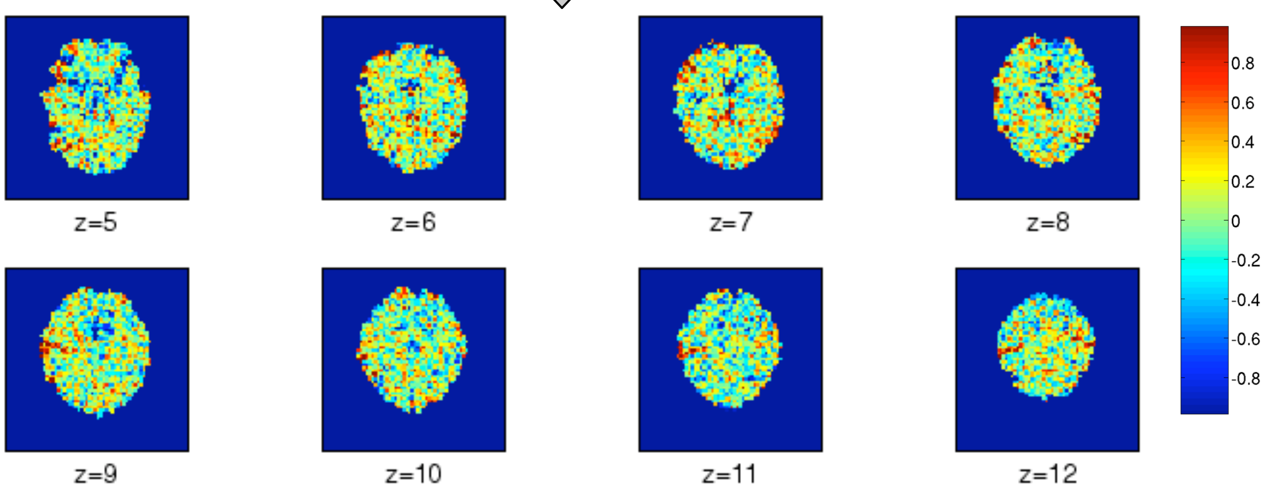
2 categories linear discriminants

It's possible to predict category using all the voxels



GNB weights
(accuracy 65%)

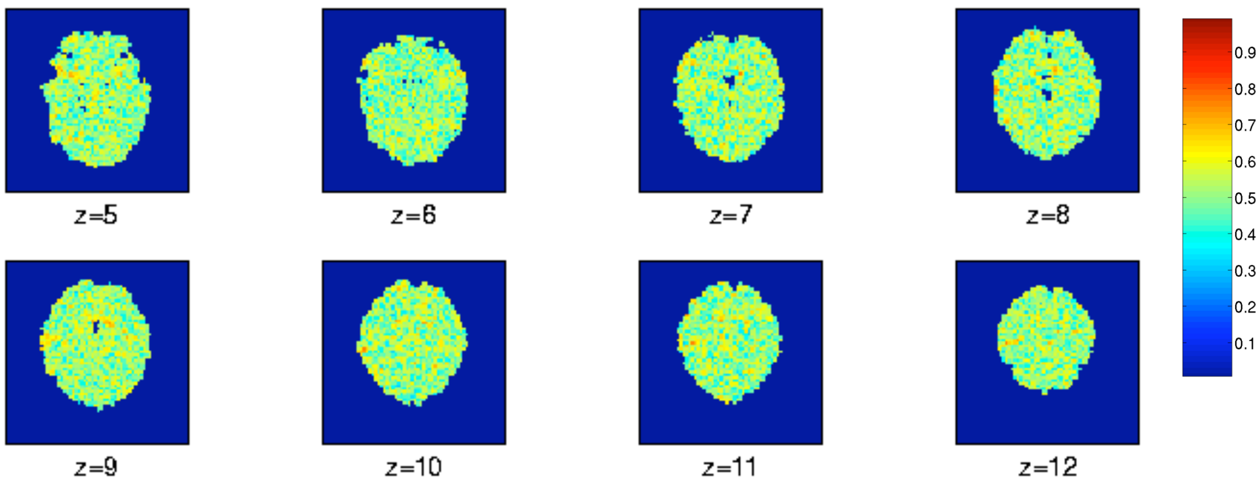
correlation 0.8



L_2 Logistic Regression weights
(accuracy 74%)

2 categories voxel accuracy maps

What is each voxel contributing?

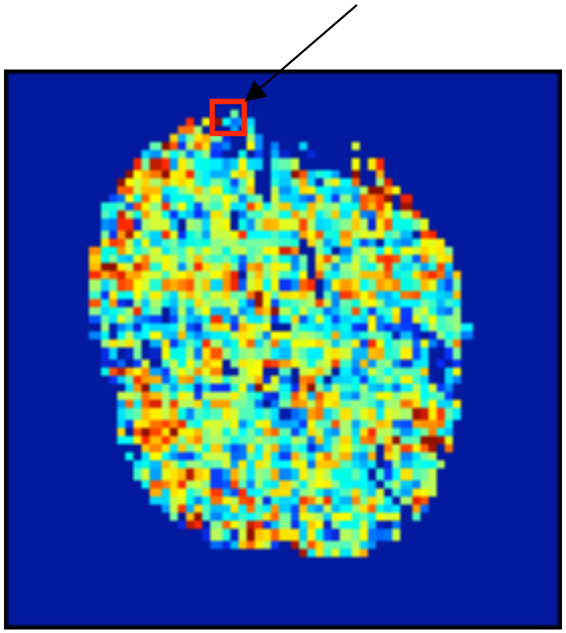


accuracy of voxelwise prediction

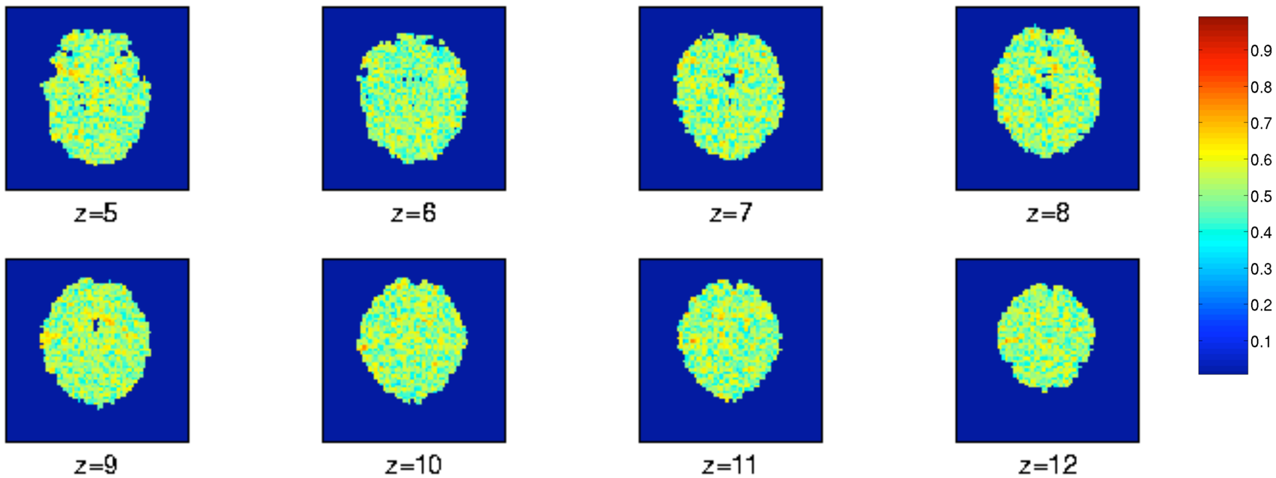
voxel searchlight

[Kriegeskorte 2006]:

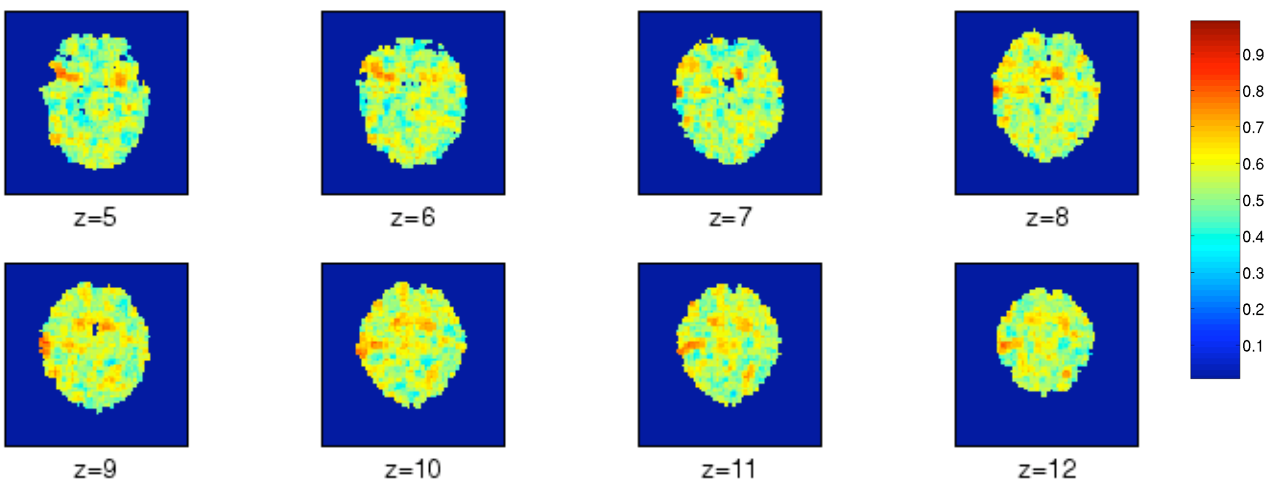
- Examine information inside a small region
- Train a classifier for each voxel together with its neighbours



2 categories voxel accuracy maps



accuracy of voxel prediction



accuracy of voxel searchlight prediction (similar in other subjects)

experiments – voxel selection

- Scoring methods for voxel selection
  - activation (different from zero in at least one class)
  - accuracy (training set cross-validation accuracy of a voxel)
  - searchlight accuracy (same but accuracy of voxel+neighbours)

experiments – voxel selection

- Scoring methods for voxel selection
  - activation (different from zero in at least one class)
  - accuracy (training set cross-validation accuracy of a voxel)
  - searchlight accuracy (same but accuracy of voxel+neighbours)

- Filter voxel selection in each fold
  - rank voxels by their score according to a method
  - pick top 10, top 20, top 40, etc

10 exemplar experiment

- subjects read concrete nouns in 2 categories
  - words are either tools or buildings
  - task:

    see a word/think about it for 3 sec., 8 sec. pause afterwards
- subjects do the same task with drawings

- Classification task: predict the exemplar
- Example:

  average 3D image middle 4 secs of a trial
- 6 examples of each exemplar

10 exemplar experiment

Peak accuracy selecting 400 voxels with 3 methods:

	GNB	Log.Reg.
all cortex voxels	23%	22%

10 exemplar experiment

Peak accuracy selecting 400 voxels with 3 methods:

	GNB	Log.Reg.
activation	70%	58%
accuracy	72%	70%
searchlight accuracy	90%	92%
all cortex voxels	23%	22%

10 exemplar experiment

Peak accuracy selecting 400 voxels with 3 methods:

	GNB	Log.Reg.	Fold Overlap
activation	70%	58%	0.09
accuracy	72%	70%	0.01
searchlight accuracy	90%	92%	0.26
all cortex voxels	23%	22%	

$$\frac{\text{\#voxels selected on all folds}}{\text{\#voxels selected on any fold}} = \text{overlap}$$

10 exemplar experiment

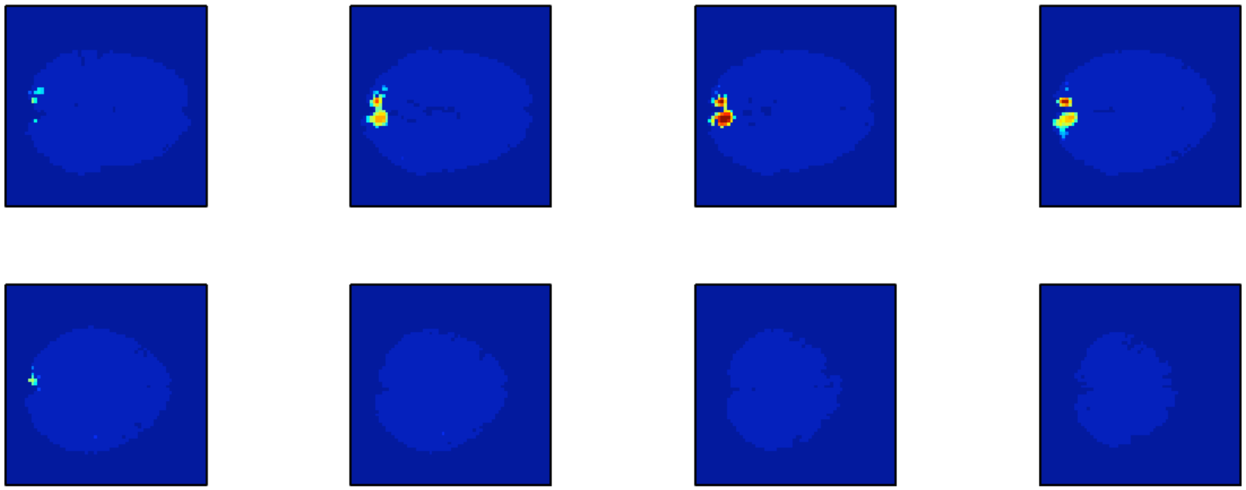
Peak accuracy selecting 400 voxels with 3 methods:

	GNB	Log.Reg.	Fold Overlap
activation	70%	58%	0.09
accuracy	72%	70%	0.01
searchlight accuracy	90%	92%	0.26
all cortex voxels	23%	22%	

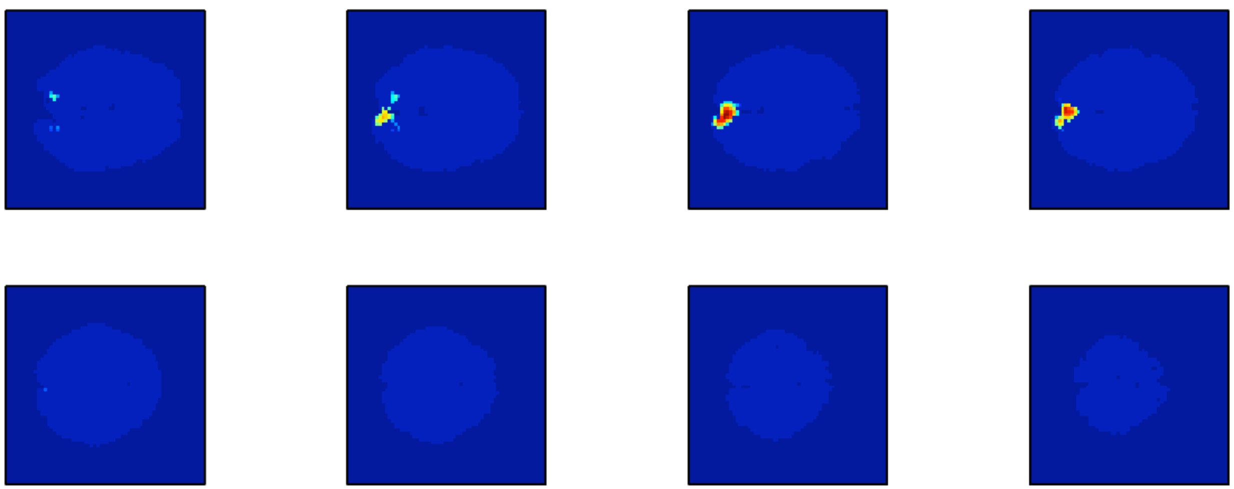
What makes searchlight accuracy better here?

10 exemplar experiment

searchlight
selected voxels
picture stimuli



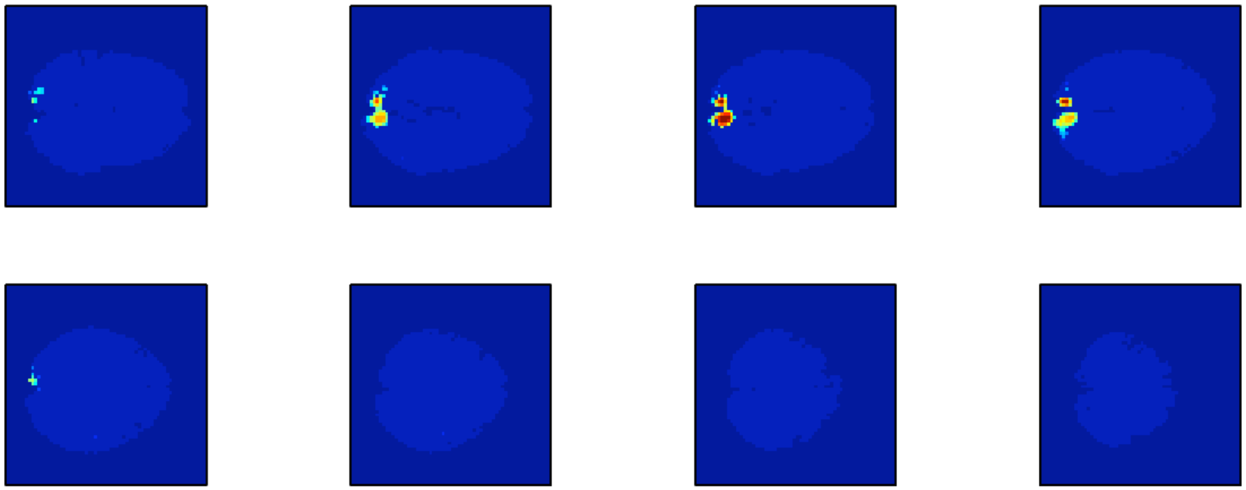
subject 1

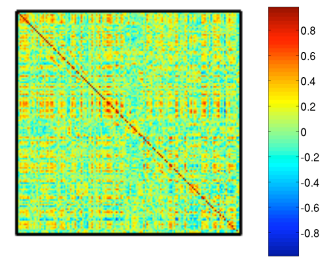


subject 2

10 exemplar experiment

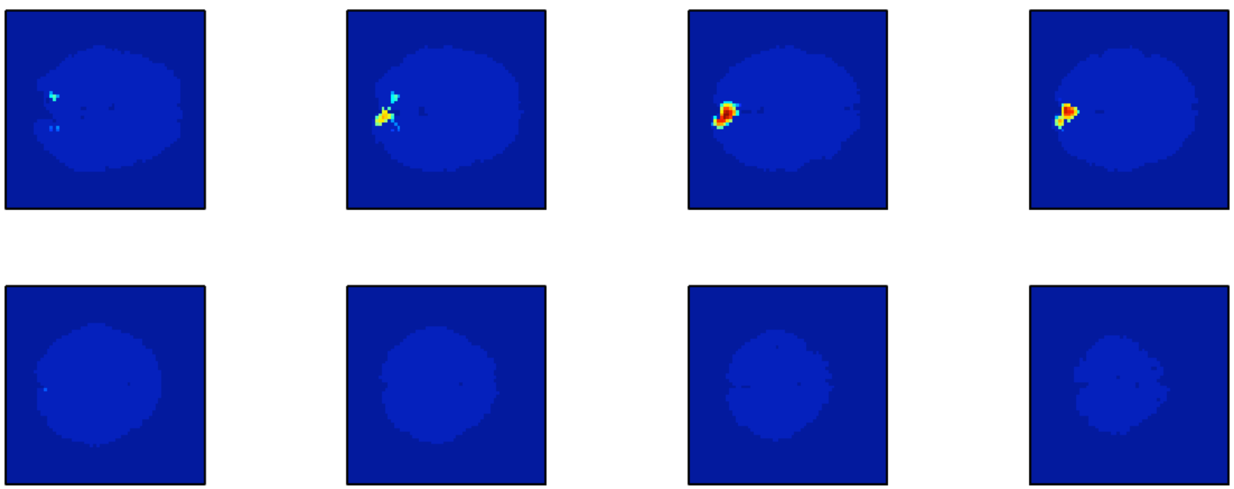
searchlight
selected voxels
picture stimuli

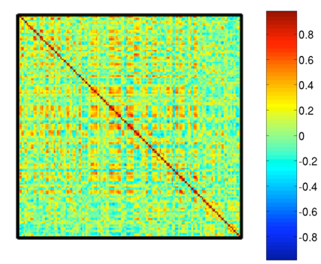




subject 1

voxel
correlation





voxel
correlation

subject 2

classifier experiment conclusions

- What should we consider?
  - interpretation depends on location/selection criteria
  - classifier regularization also plays a role
  - information is redundant
  - information is local

classifier experiment conclusions

- What should we consider?
  - interpretation depends on location/selection criteria
  - classifier regularization also plays a role
  - information is redundant
  - information is local
- What should we care about?
  - prediction accuracy
  - describing what was learnt intelligibly
    - location
    - voxel behaviour reduced to a few classes
    - voxel groupings/data abstraction
  - reproducibility [Strother 2002]
  - consistency with prior knowledge (mostly location)

What is to be done?

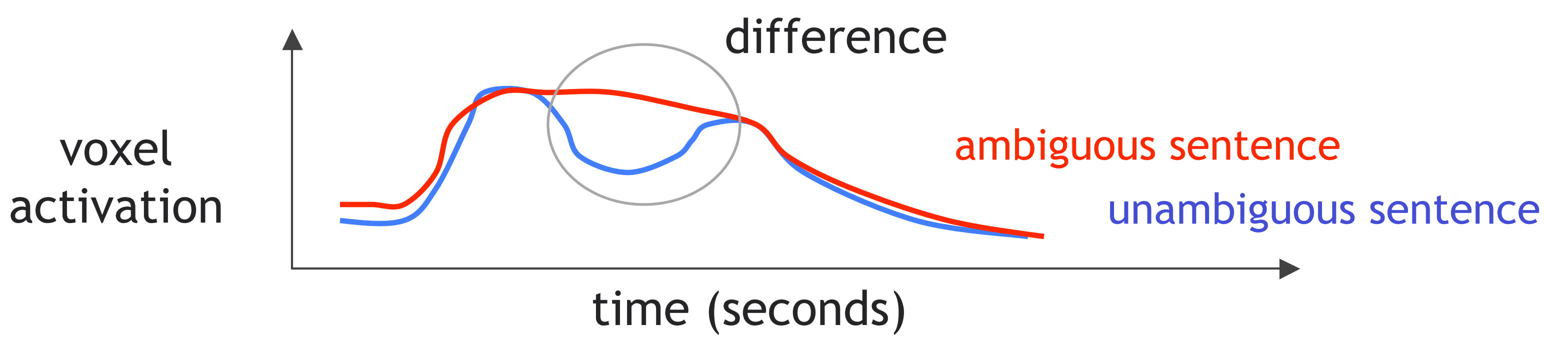
- Get more data into play
- Model time or other parts of fMRI process
- Predictions other than stimuli
- Learn data abstractions
- Use prior knowledge

What is to be done?

- Get more data into play
  - use multiple subjects from the same study
    - structural normalization (brain morph)
    - functional normalization (activity morph)
    - models have subject specific/subject independent parts
  - use the same subject in multiple studies
  - transfer/multitask learning

What is to be done?

- Model time or other parts of fMRI process
  - use voxels at a given time in a trial
  - model trial response and learn classifiers for that



What is to be done?

- Predictions other than stimuli
  - subjective mental states
  - decisions
  - subconscious processing
  - group membership (diagnosis)
  - behavioural measures

What is to be done?

- Use prior knowledge/hypotheses
  - brain areas/connections involved
  - spatial locality
    - neighbouring voxels have similar activity
    - neighbouring voxels classifier weights have similar magnitude
    - groups of voxels are acting together “interestingly”

If weight 0 + weight1 x voxel 1 + weight2 x voxel 2 + … + … + weight n x voxel n > 0 tools
otherwise buildings

  - cognitive models

What is to be done?

- Use prior knowledge/hypotheses
  - brain areas/connections involved
  - spatial locality
    - neighbouring voxels have similar activity
    - neighbouring voxels classifier weights have similar magnitude
    - groups of voxels are acting together "interestingly"

If weight 0 + weight1 x voxel 1 + weight2 x voxel 2 + … + … + weight n x voxel n > 0 tools
otherwise buildings

  - cognitive models

What is to be done?

- Use prior knowledge/hypotheses
  - brain areas/connections involved
  - spatial locality
    - neighbouring voxels have similar activity
    - neighbouring voxels classifier weights have similar magnitude
    - groups of voxels are acting together “interestingly”

If weight 0 + weight1 x voxel 1 + weight2 x voxel 2 + … + … + weight n x voxel n > 0 tools

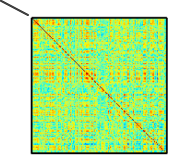
otherwise buildings

  - cognitive models

What is to be done?

- Use prior knowledge/hypotheses
  - brain areas/connections involved
  - spatial locality
    - neighbouring voxels have similar activity
    - neighbouring voxels classifier weights have similar magnitude
    - groups of voxels are acting together "interestingly"

If weight 0 + weight1 x voxel 1 + weight2 x voxel 2 + … + … + weight n x voxel n > 0 tools
otherwise buildings

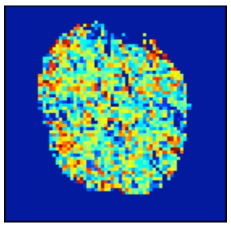
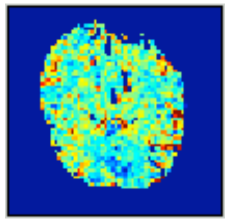
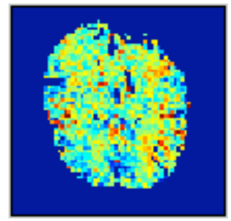
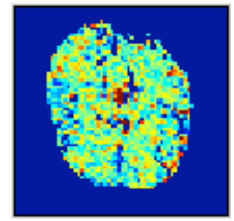
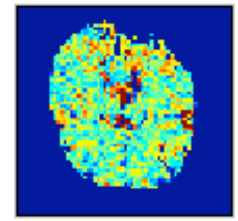


  - cognitive models

What is to be done?

- Learn and use data abstractions
  - blobs/clusters
  - interacting groups
  - brain-wide components
  - subject specific/shared across subjects
  - non linear classifiers in terms of these?

low-dimensional spatial decompositions

 = a  + b  + c  + d 

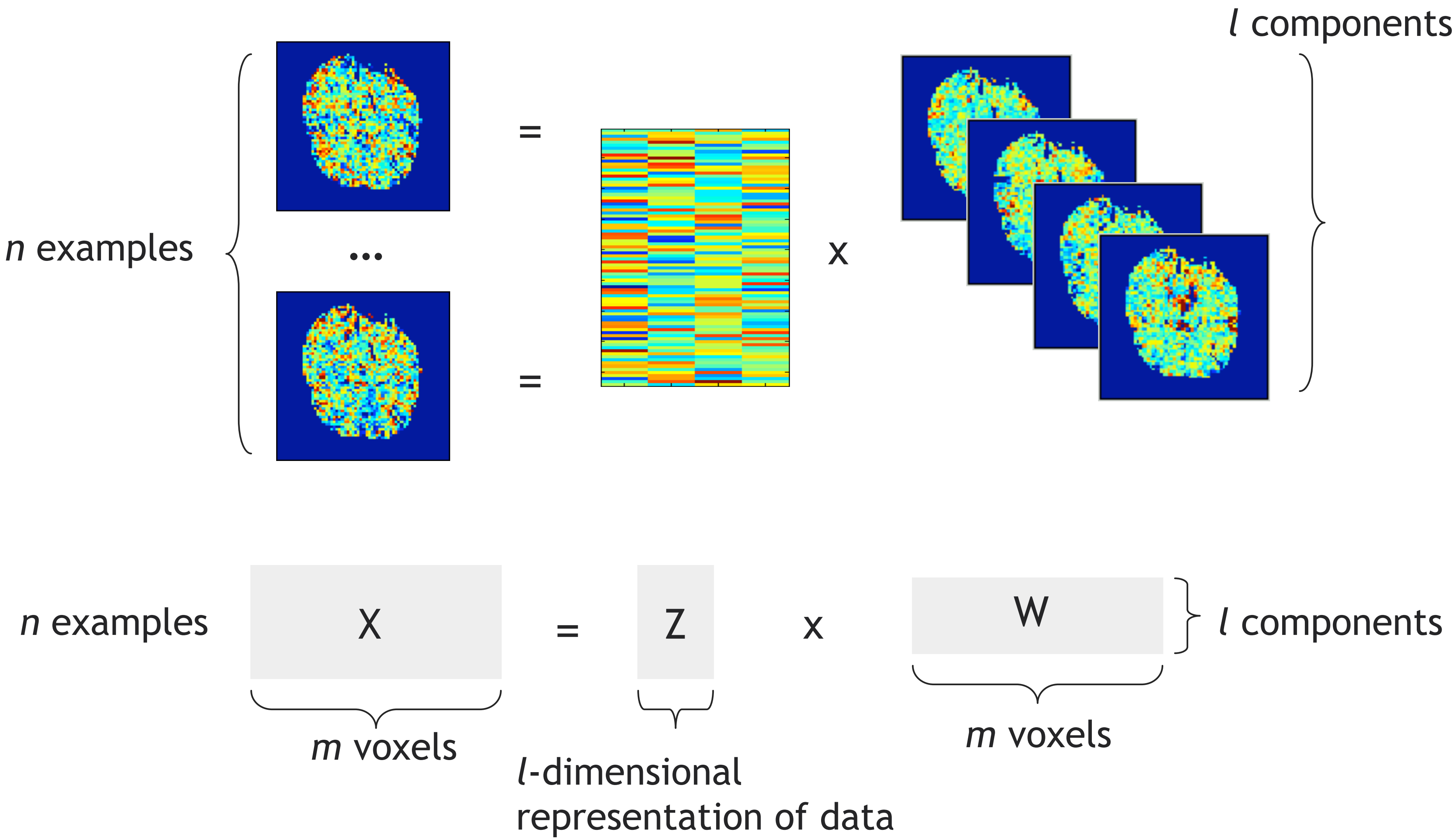
example

components or eigenimages

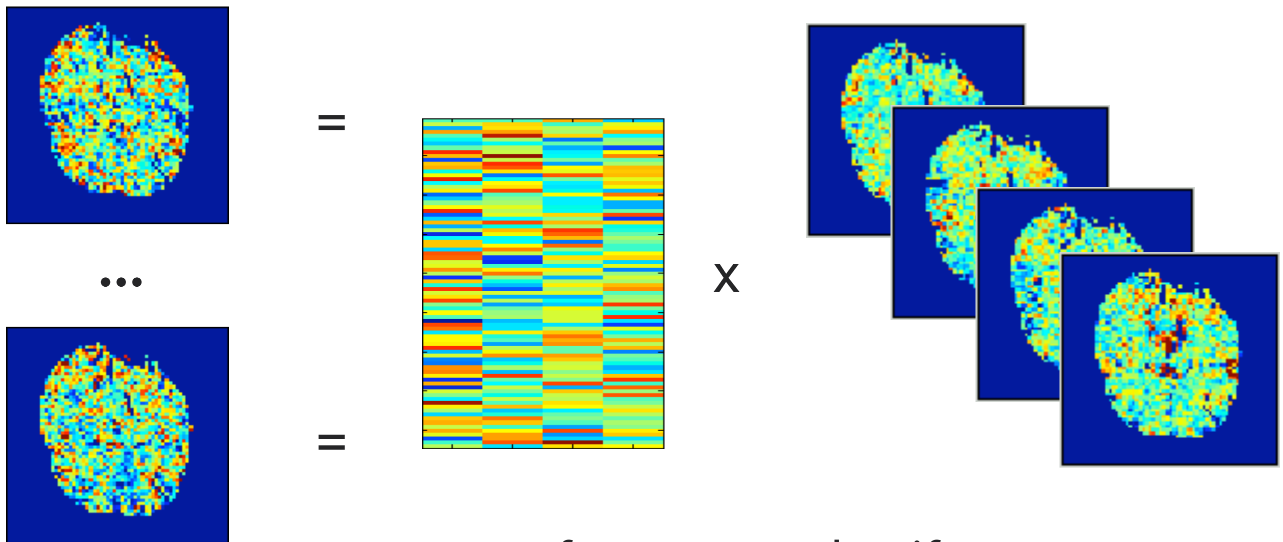
(a,b,c,d)
is a low-dimensional
representation of
the example

in a basis of components

low-dimensional spatial decompositions



combining decompositions with classifiers



new features to classify
from with linear discriminant Θ

$$\underbrace{\underbrace{X}_{m \text{ voxels}}}_{n \text{ examples}} = \underbrace{Z}_{l\text{-dimensional representation of data}} \times \underbrace{\underbrace{W}_{m \text{ voxels}}}_{l \text{ components}}$$

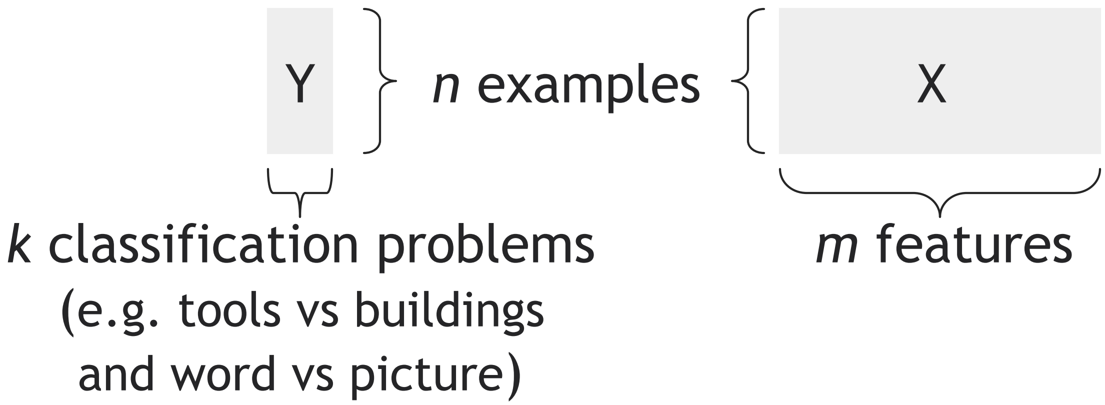
support vector decomposition machine (SVDM)

Learning a linear SVM based on a low-dimensional representation

Learning an informative low-dimensional representation

[Pereira&Gordon 2006]

SVDM notation



SVDM notation

$$\underbrace{Y}_{k \text{ classification problems}} \Big\} \; n \text{ examples} \; \Big\{ \underbrace{X}_{m \text{ features}}$$

Predictions

Learnt

$$\hat{X} = \overbrace{Z}^{l \text{ components}} \times \overbrace{W}^{m \text{ features}} \Big\} \; l \text{ components}$$

$$\hat{Y} = \text{sign} \left[Z \times \underbrace{\Theta}_{k \text{ classification problems}} \right]$$

SVDM work in progress

- Multi-class
  - Learn components shared by subsets of the classes
- Multi-subject

$$\begin{bmatrix} X_1 & X_2 \end{bmatrix} = Z \begin{bmatrix} W_1 & W_2 \end{bmatrix}$$

- Constraints
  - classifier regularization
  - component smoothness/sparsity
  - voxel behaviour (e.g. active in few classes)
  - hypothesis-driven component sharing

What is to be done?

- Get more data into play
- Model time or other parts of fMRI process
- Predictions other than stimuli
- Learn data abstractions
- Use prior knowledge

- Doing well is much more than being accurate
- No science without hypotheses

thank you!

Questions?

*No classifiers were harmed in producing this talk. Some grad students may have been.