Statistical Pattern Recognition with fMRI

Heini Saarimäki
University Instructor (Statistical Methods in Psychology)
Human Information Processing Laboratory
Tampere University

Outline

WHAT: Nuts and bolts of pattern recognition with fMRI

WHY: When and why to use pattern recognition methods

HOW: Overview of the workflow

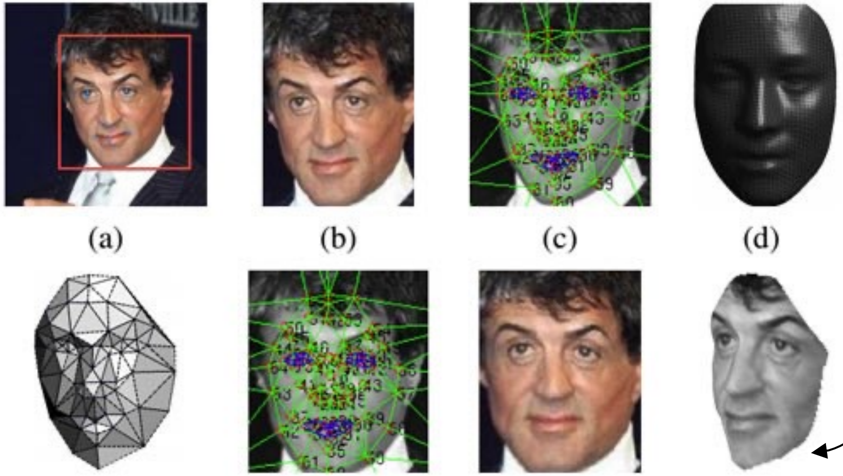
What

Statistical pattern recognition

Features and their configuration make the Sylvester pattern

A sample

”Sylvester”



(a) (b) (c) (d)

(e) (f) (g) (h)

A different type of sample

Statistical pattern recognition with fMRI

Multivoxel pattern analysis (MVPA)

(Supervised) machine learning

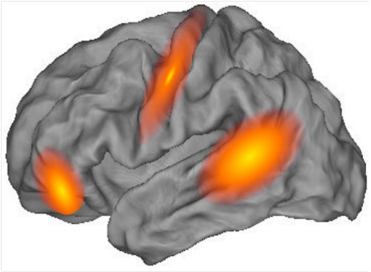
Decoding

Classification

…

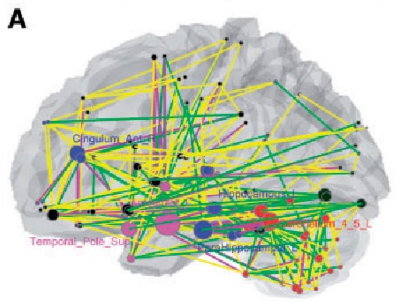
Brain activity pattern

"Fear"



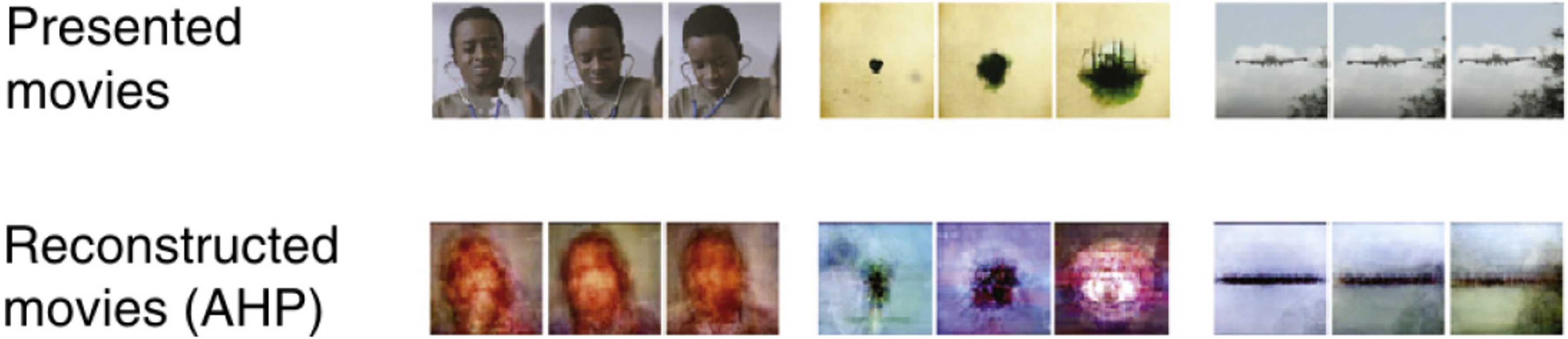
Functional connectivity pattern

"Depression"



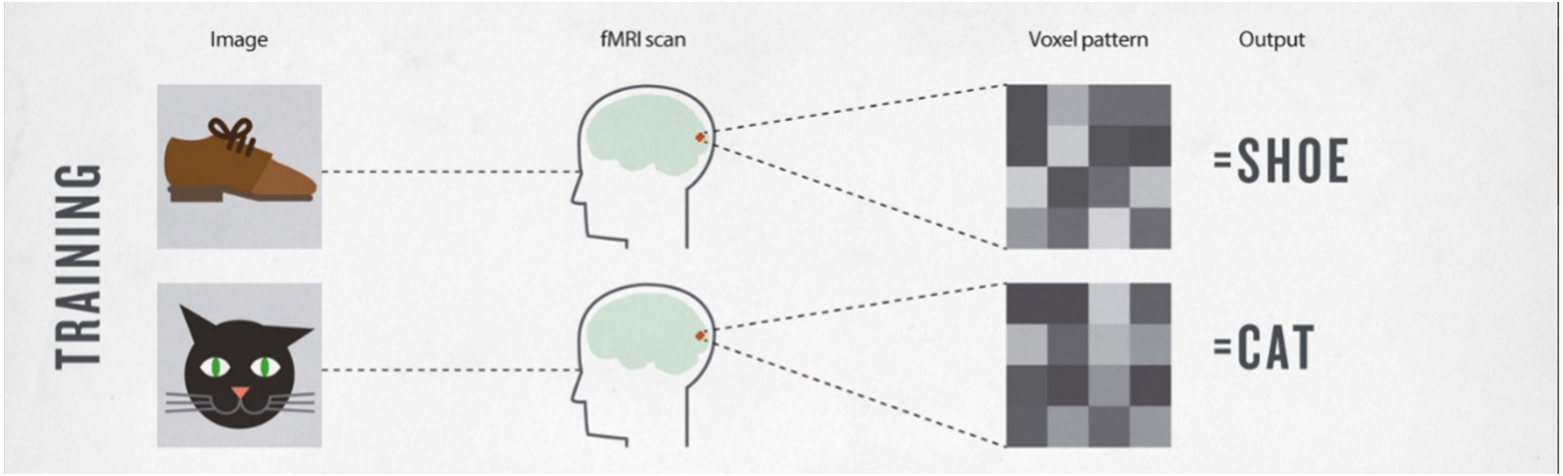
Zeng et al. (2012) *Brain*

Decoding movies from visual cortex



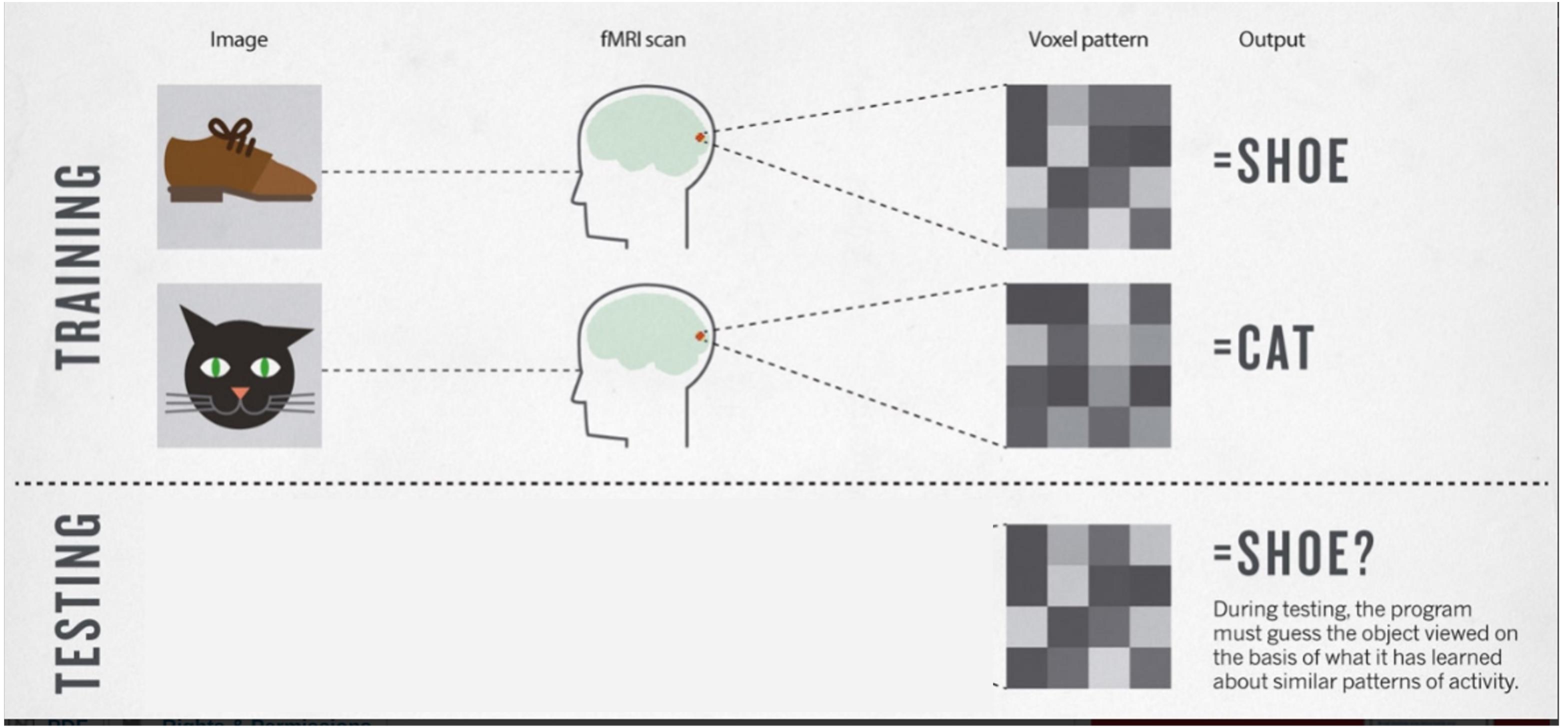
Nishimoto et al. (2011) *Curr Biol*

Simplified framework:



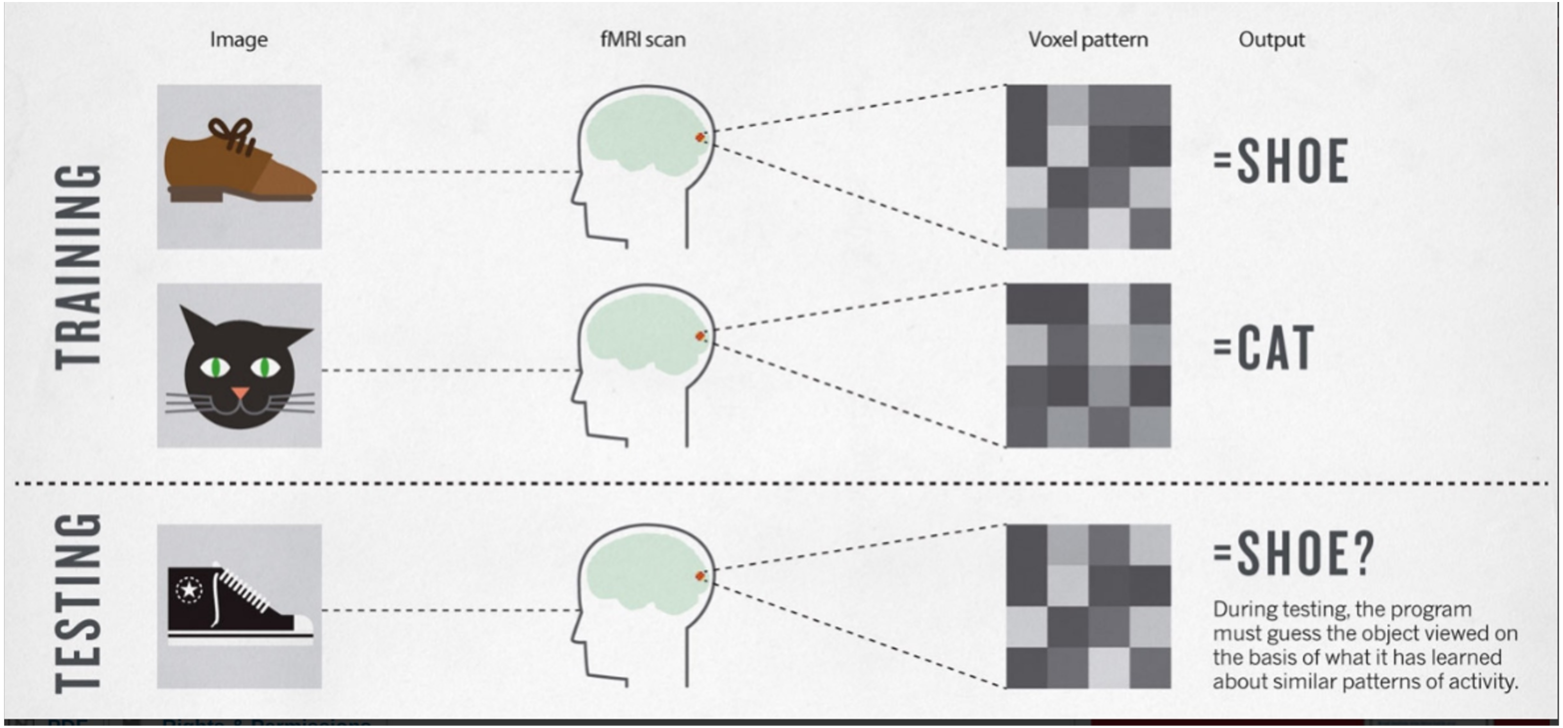
e.g. Haxby et al. (2001) *Science*
Figure credits: Smith (2013) *Nature*

Simplified framework:



e.g. Haxby et al. (2001) *Science*
Figure credits: Smith (2013) *Nature*

Simplified framework:



e.g. Haxby et al. (2001) *Science*
Figure credits: Smith (2013) *Nature*

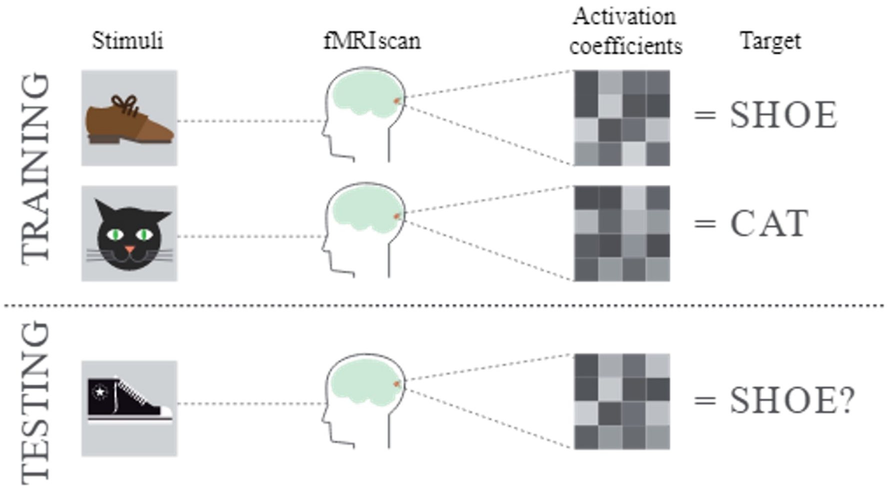
Applications

Can we distinguish the brain activity underlying some mental states?

Can we distinguish patient groups by looking at their brain?

Medical Diagnosis

Brain-Computer-Interface



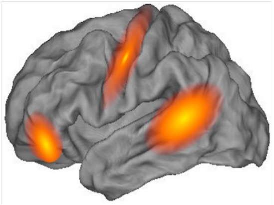
Why

Why pattern recognition?

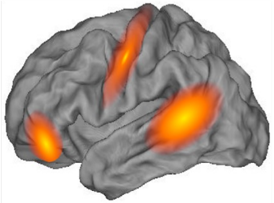
Multivariate, thus has higher sensitivity when compared to univariate analyses
→ Can detect pattern differences when activation overlaps

Univariate

Fear



Disgust

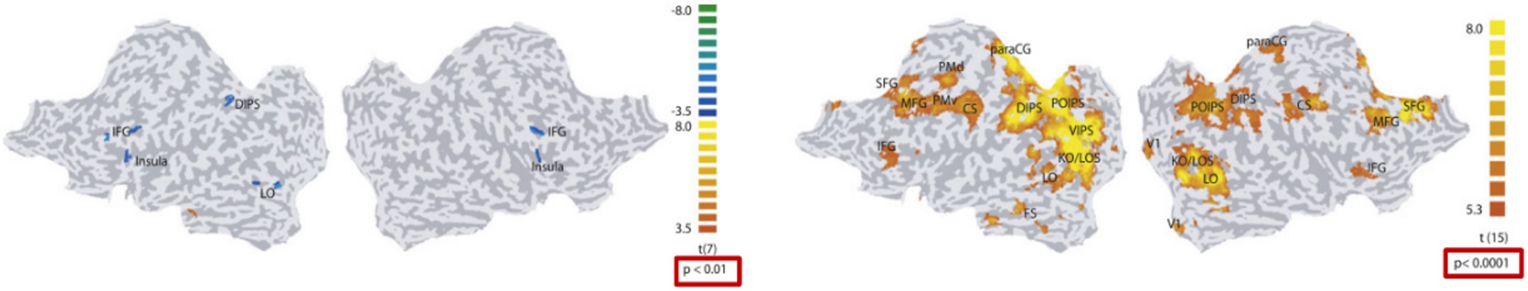


Why pattern recognition?

Multivariate, thus has higher sensitivity when compared to univariate analyses
→ Can detect pattern differences when activation overlaps

Example: Representation of perceptual choices

univariate multivariate

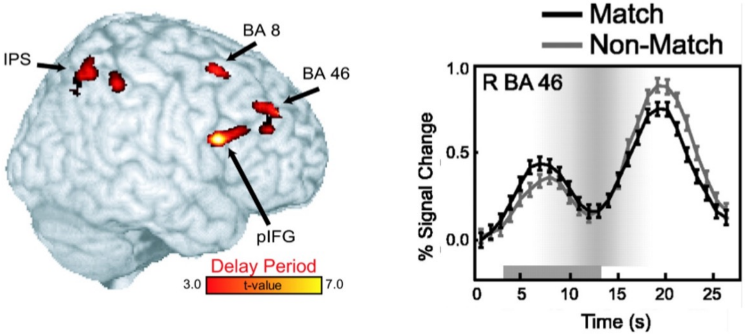


Li et al. (2009) *Neuron*

Why pattern recognition?

We can study representational content in a brain region rather than general activation

Activity: Tells us about general involvement in cognitive function (e.g. working memory)

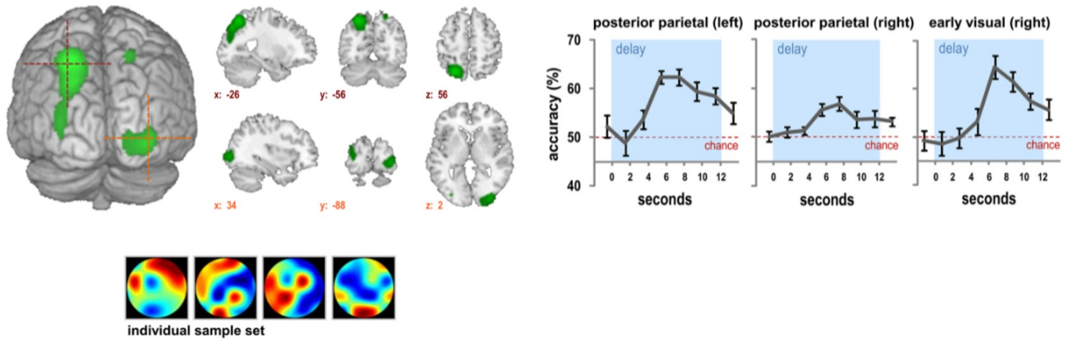


Curtis et al. (2004) *J Neurosci*

Why pattern recognition?

We can study representational content in a brain region rather than general activation

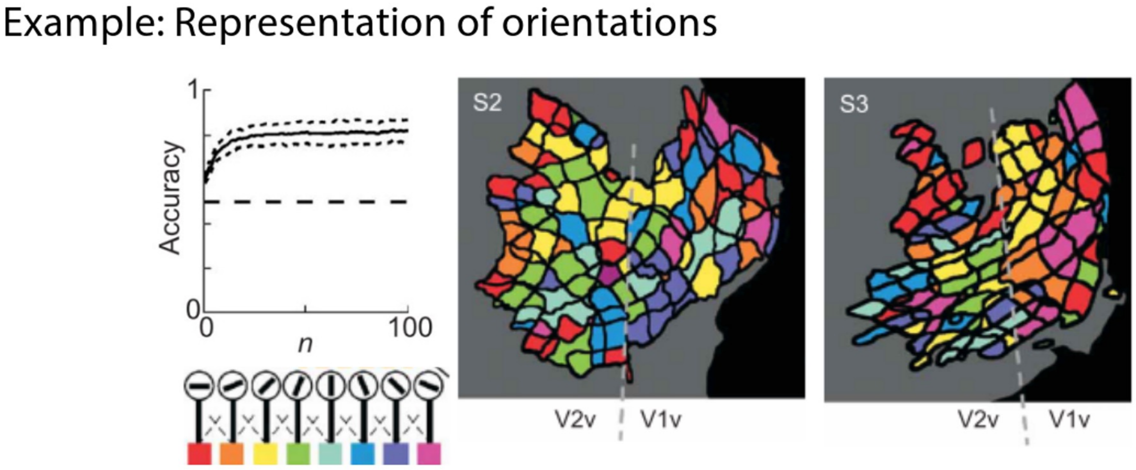
Information: Tells us about representational content (e.g. memory trace of A vs. memory trace of B)



Christophel et al. (2012) *J Neurosci*

Why pattern recognition?

We can study representational content in a brain region rather than general activation



Kamitani & Tong (2005) *Nat Neurosci*

Why pattern recognition?

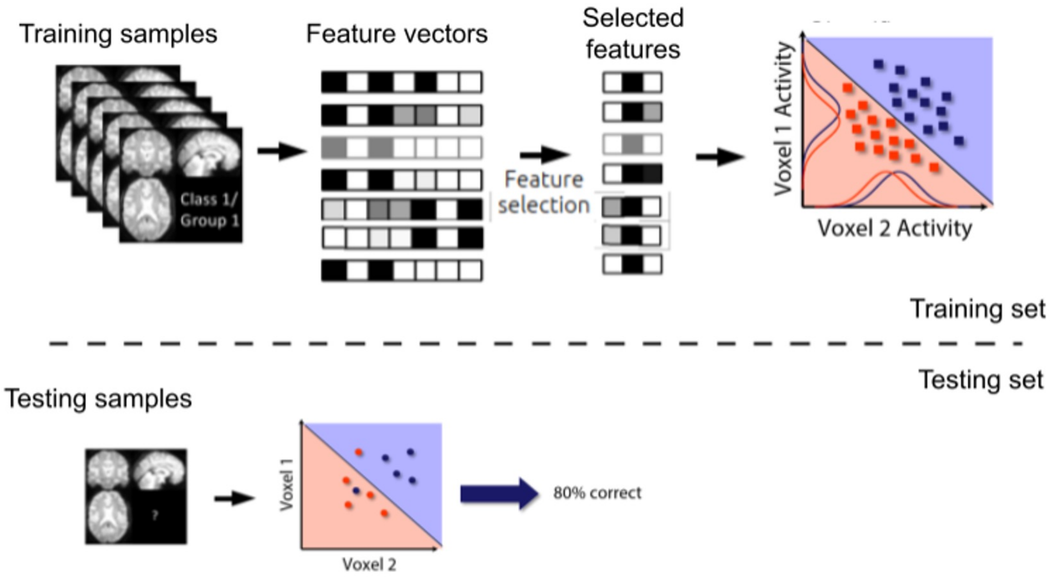
Methods readily available from machine learning

The principle is relatively easy to understand

Has become a standard method, complementary to univariate analyses

How

Overview of the workflow



Design and data acquisition

Preprocessing

Type of analysis

Feature selection

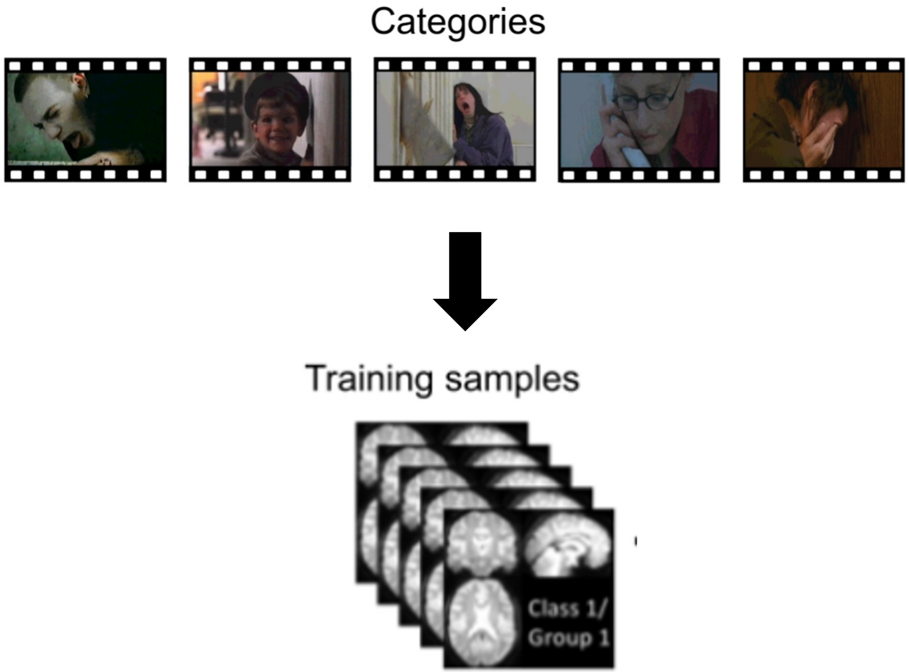
Classification

Statistical analyses

Step 1: Design and data acquisition

Goal: sample brain activity associated with different categories

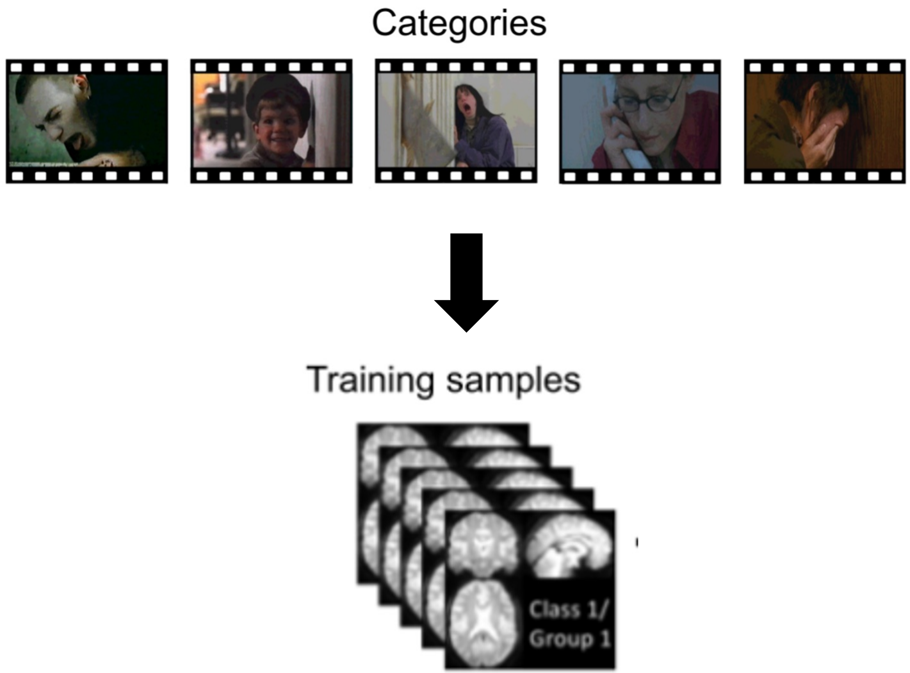
Samples could be EPI volumes, beta volumes, connectivity matrices, ...



Step 1: Design and data acquisition

GLM design principles apply

Analysis will be easier with
- balanced number of samples per category (in each run)
- careful randomization of categories (fMRI autocorrelation issues)



Step 2: Preprocessing

Minimal preprocessing

Within-subject analysis: data in native space
Between-subject analyses: data in MNI space

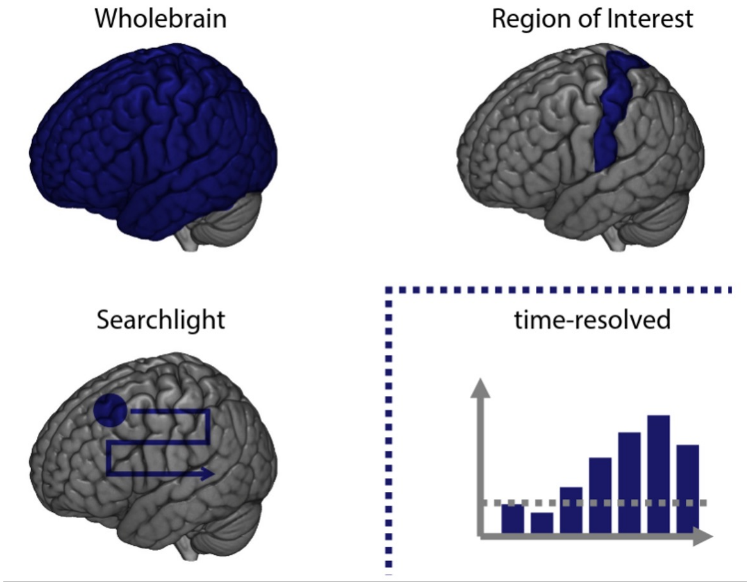
No smoothing

Normalization of signal intensities

Step 3: Type of analysis

Select the spatial and temporal range of interest.

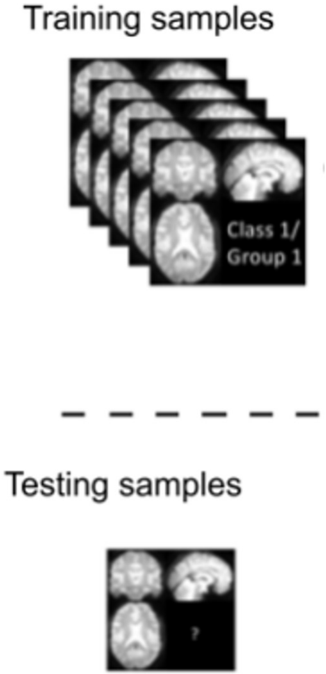
- Spatial resolution: wholebrain (global patterns), ROI or searchlight (local patterns)
- Time resolution: beta maps, single time points, ...



Step 3: Type of analysis

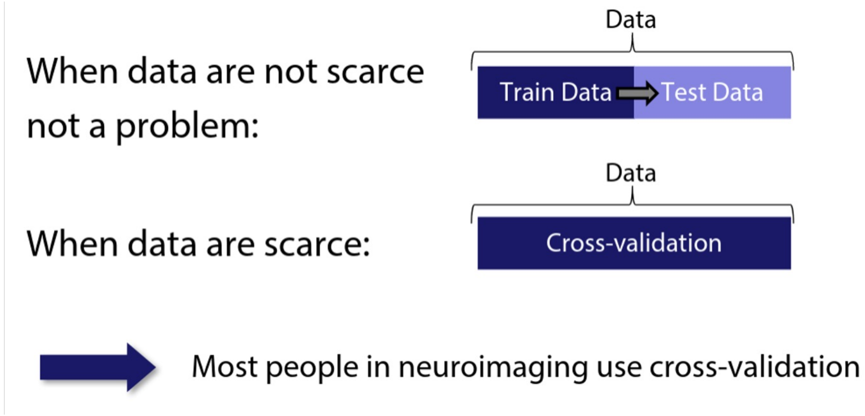
Split the data to training and testing sets.

- Goal: avoid peeking / overfitting.



Step 3: Type of analysis

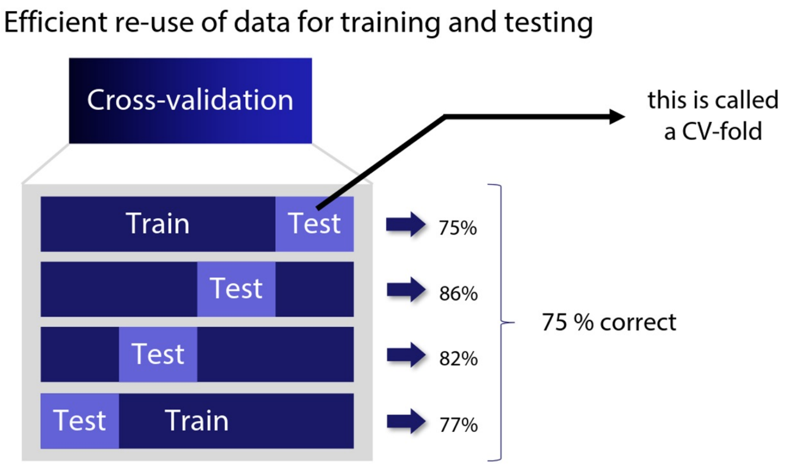
Challenge: we need to both…
i) maximize size of training data for better model fit
ii) maximize size of test data for precise generalization estimate



Step 3: Type of analysis

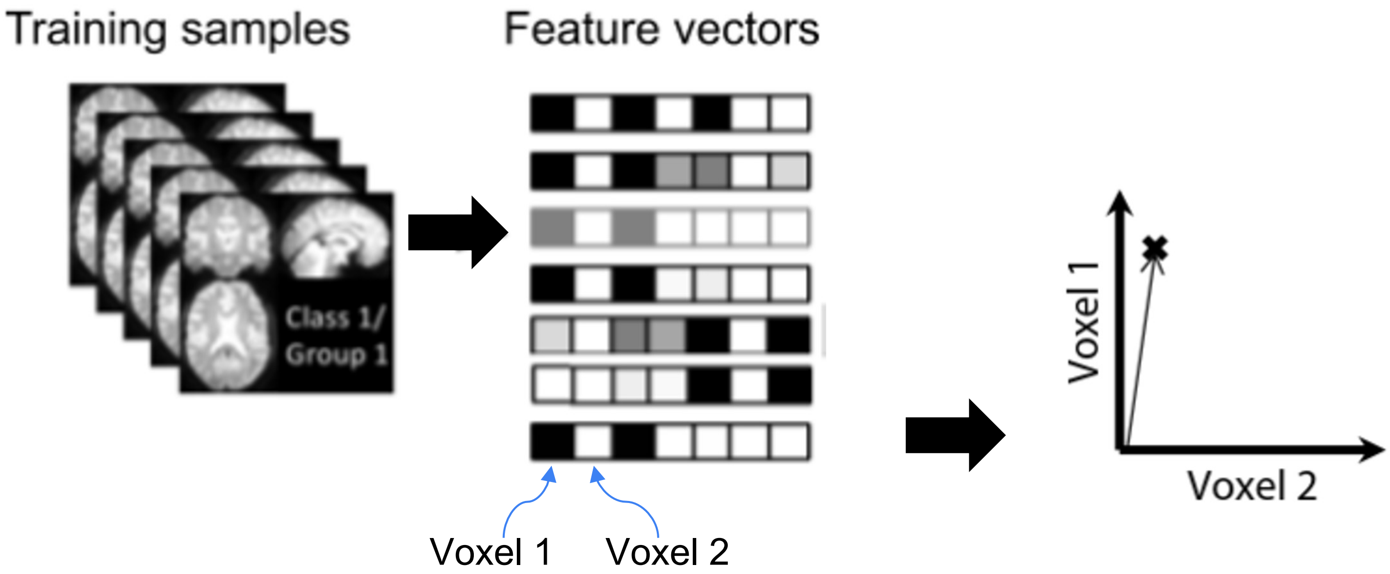
Solution: cross-validation

Cross-validation schemes e.g.
Leave-one-run-out (LORO)
Leave-one-subject-out (LOSO)



What do we have so far?

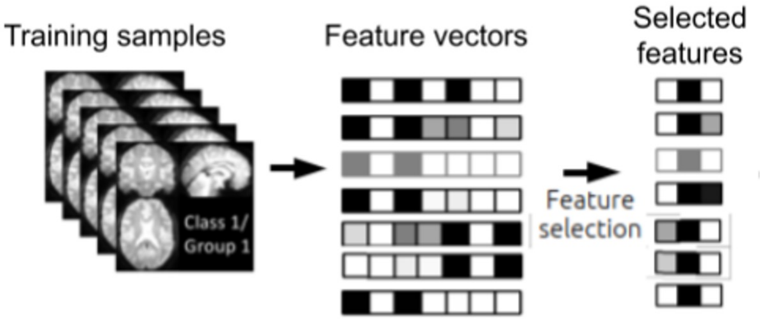
Feature = a measured variable used for classification, e.g. activity for each voxel
Pattern = a point in p-dimensional space (p = number of features)



Step 4: Feature selection

Challenge:
Too few samples, too many dimensions

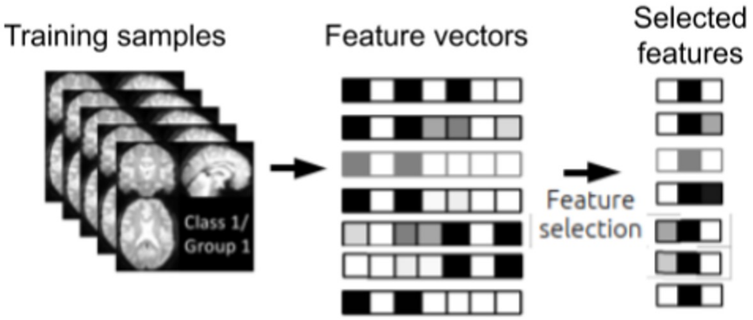
Solution:
Reduce the dimensionality by removing e.g. “uninformative” voxels



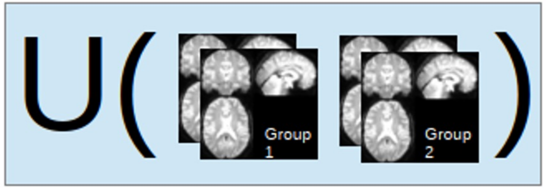
Step 4: Feature selection

Challenge:
Too few samples, too many dimensions

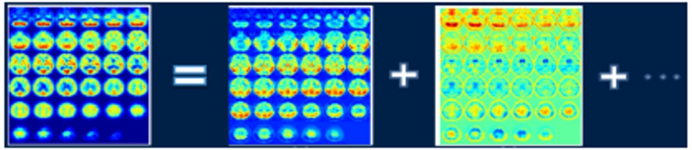
Solution:
Reduce the dimensionality by removing e.g. “uninformative” voxels



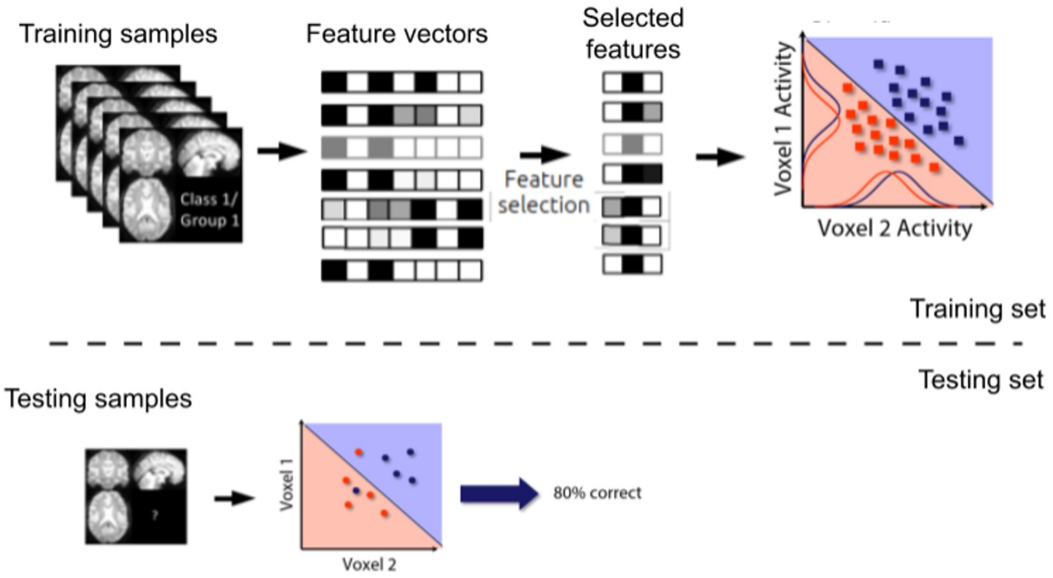
ANOVA



PCA



Ready to train and test the classifier!



Design and data acquisition

Preprocessing

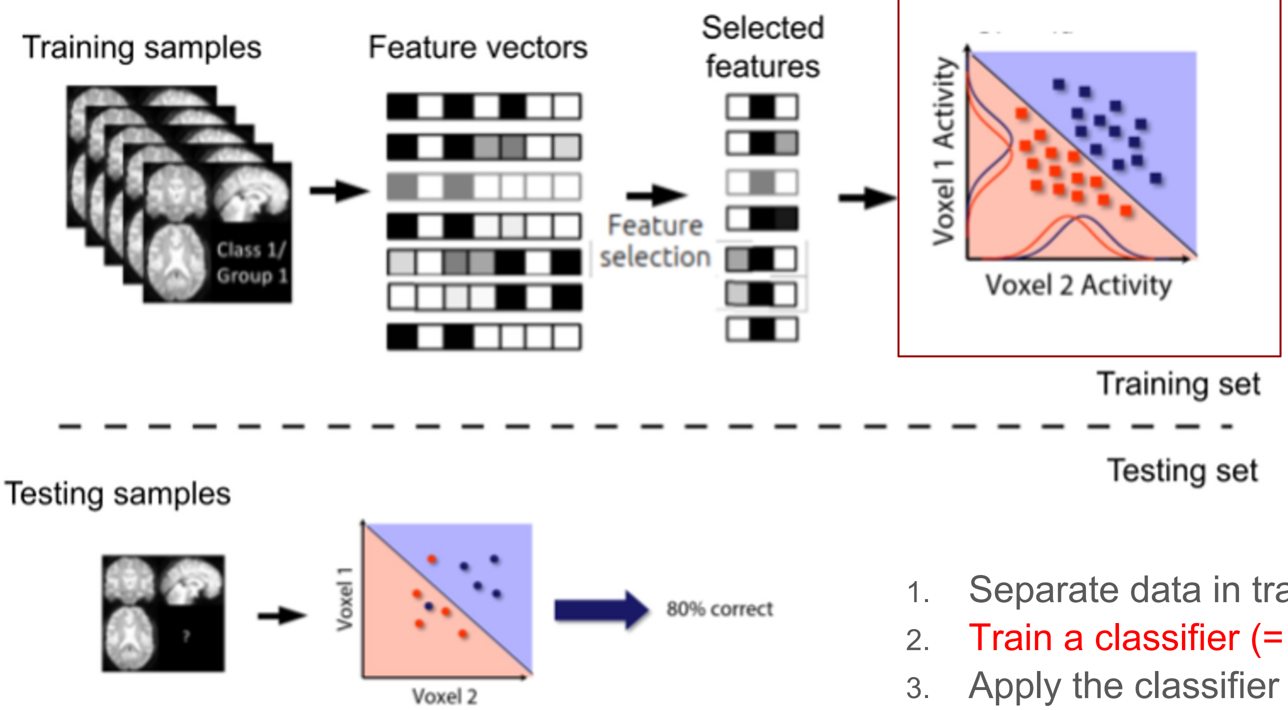
Type of analysis

Feature selection

Classification

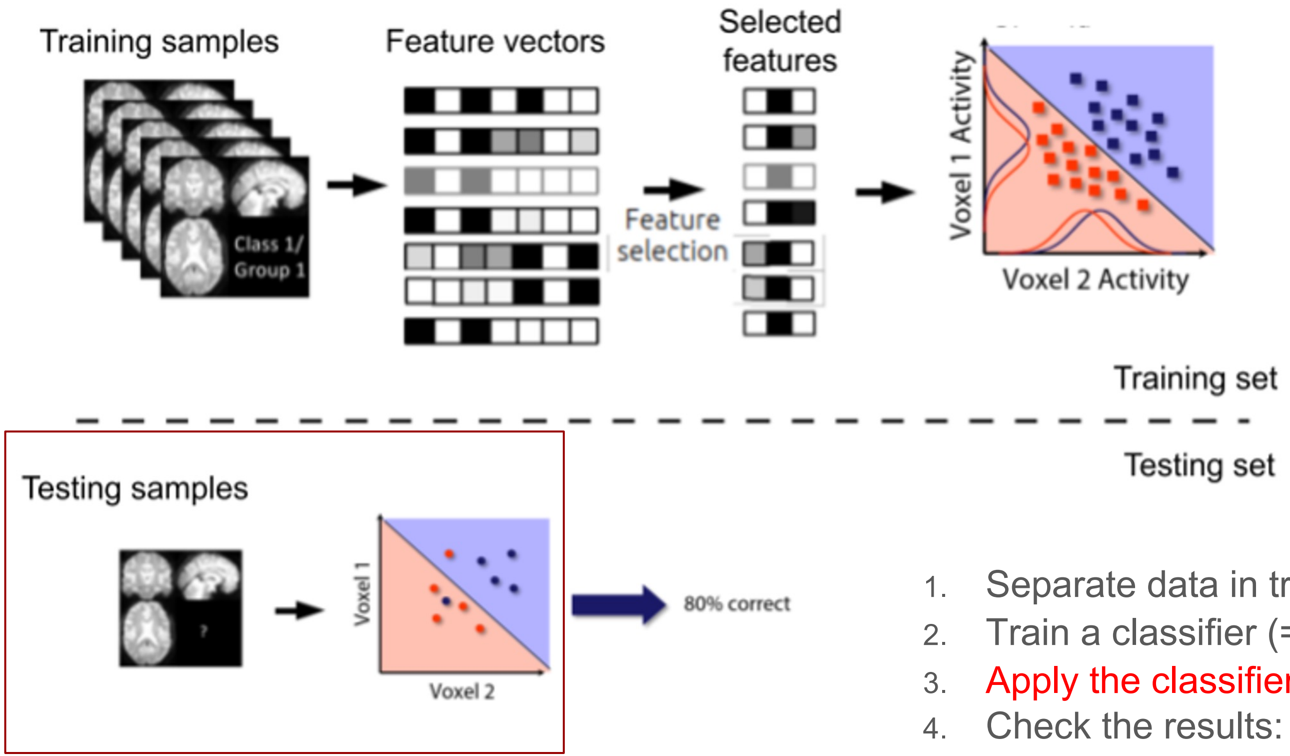
Statistical analyses

Step 5: Classification



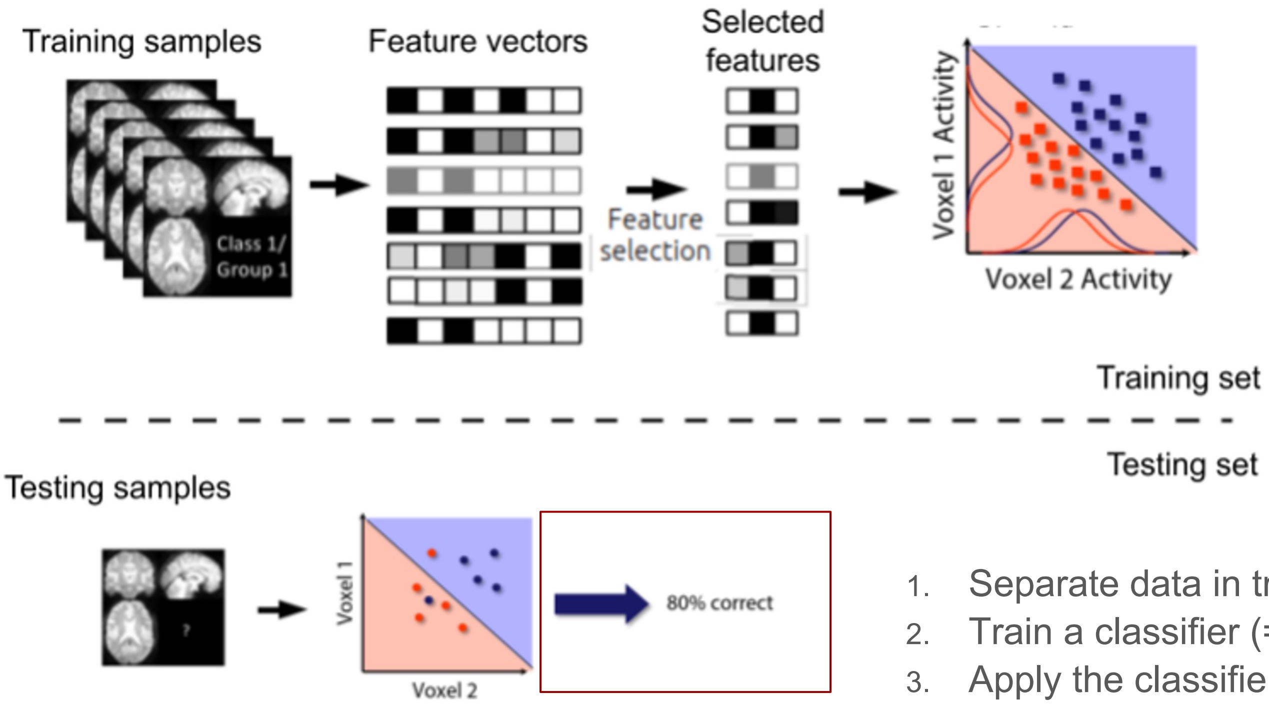
1. Separate data in training and test data
2. Train a classifier (= find a good separating line*)
3. Apply the classifier (= the line) to test data
4. Check the results: accuracy of predicting test data (=% correct prediction)

Step 5: Classification



1. Separate data in training and test data
2. Train a classifier (= find a good separating line*)
3. Apply the classifier (= the line) to test data
4. Check the results: accuracy of predicting test data (=% correct prediction)

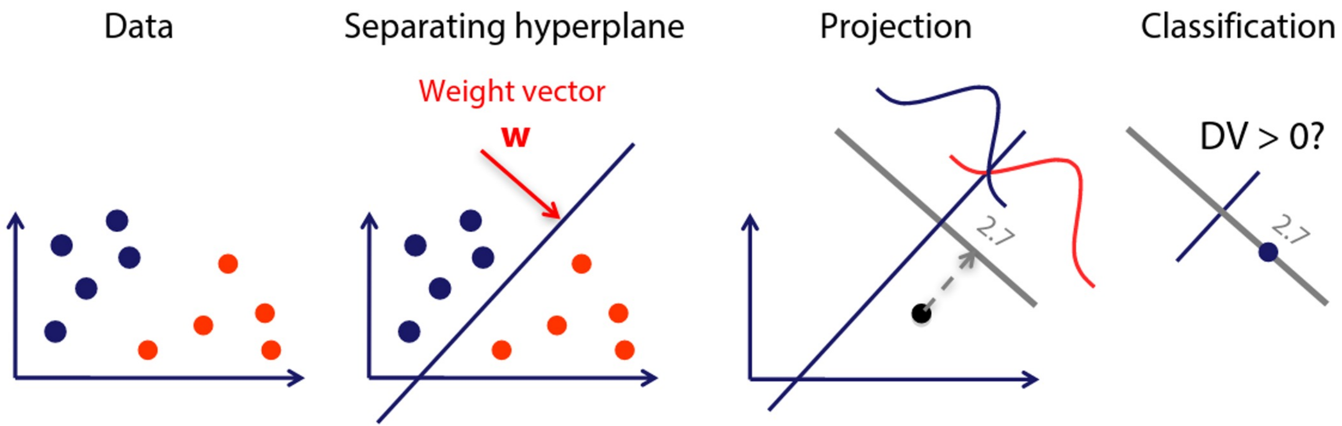
Step 5: Classification



1. Separate data in training and test data
2. Train a classifier (= find a good separating line)
3. Apply the classifier (= the line) to test data
4. Check the results: accuracy of predicting test data (=% correct prediction)

Step 5: Classification

Classifier = A function that separates feature space

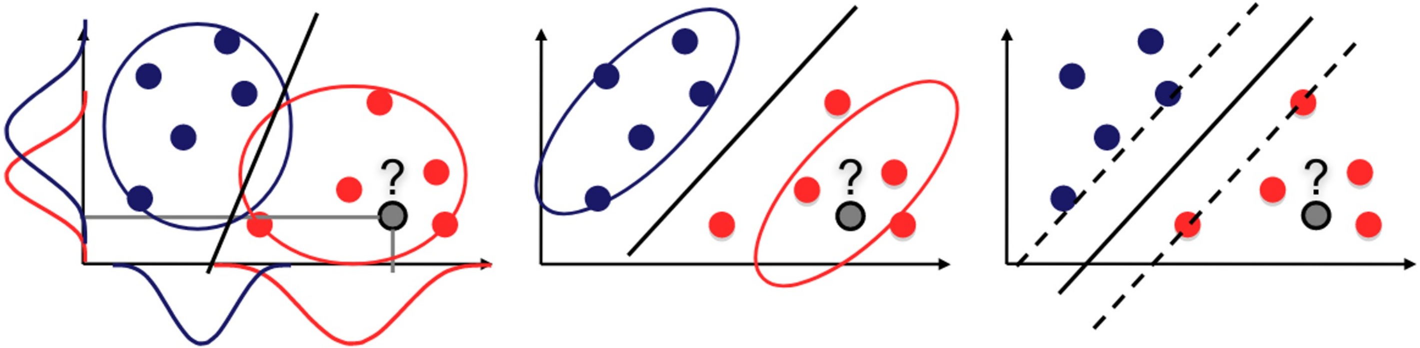


Weights are trained during classifier training

$$f(x) = w^T x + b$$

Step 5: Classification - linear classifiers

Gaussian Naïve Bayes | Linear Discriminant Analysis | Support Vector Machine

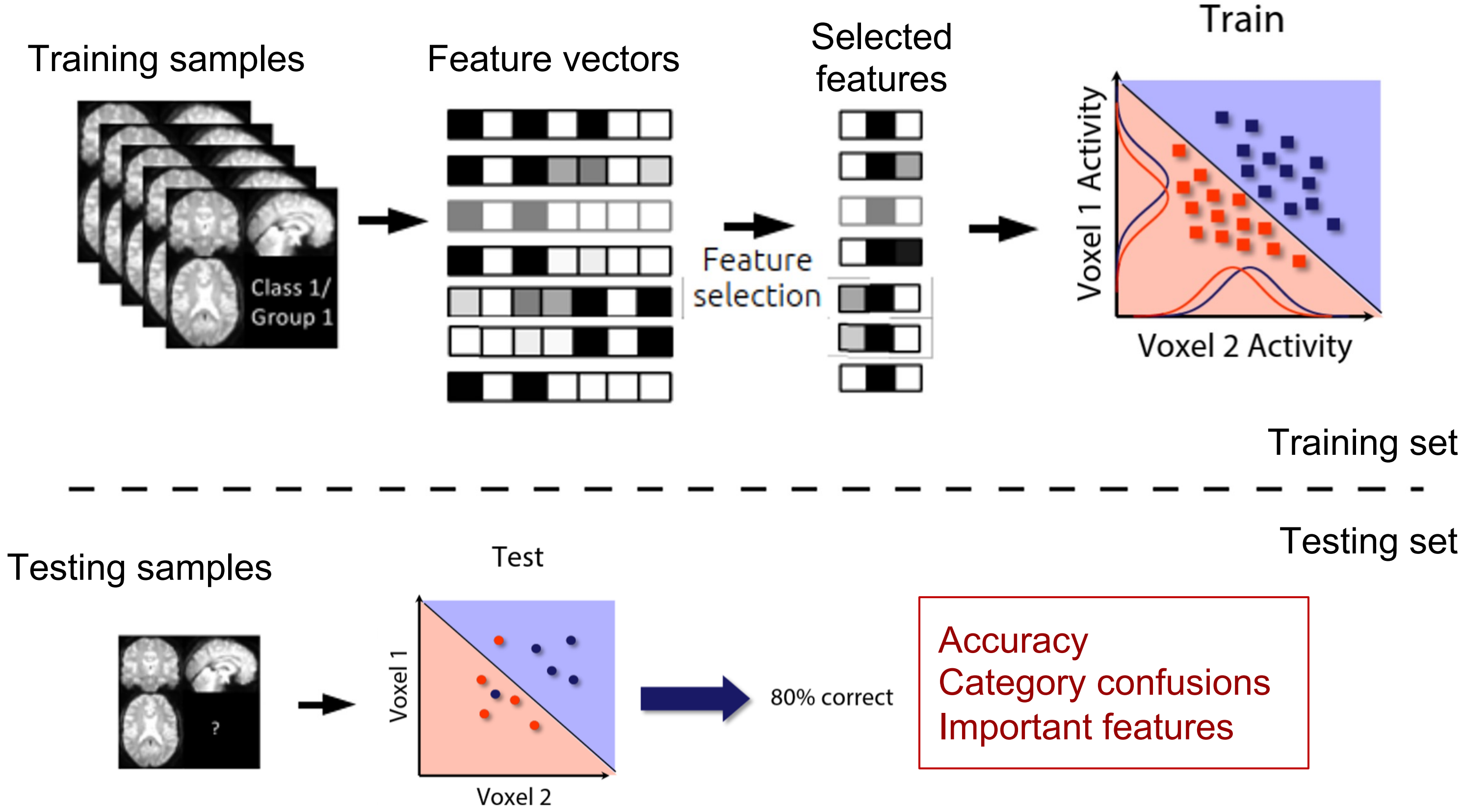


Ignores covariance between voxels

Considers covariance between voxels

Maximizes margin (distance between closest points of different classes)

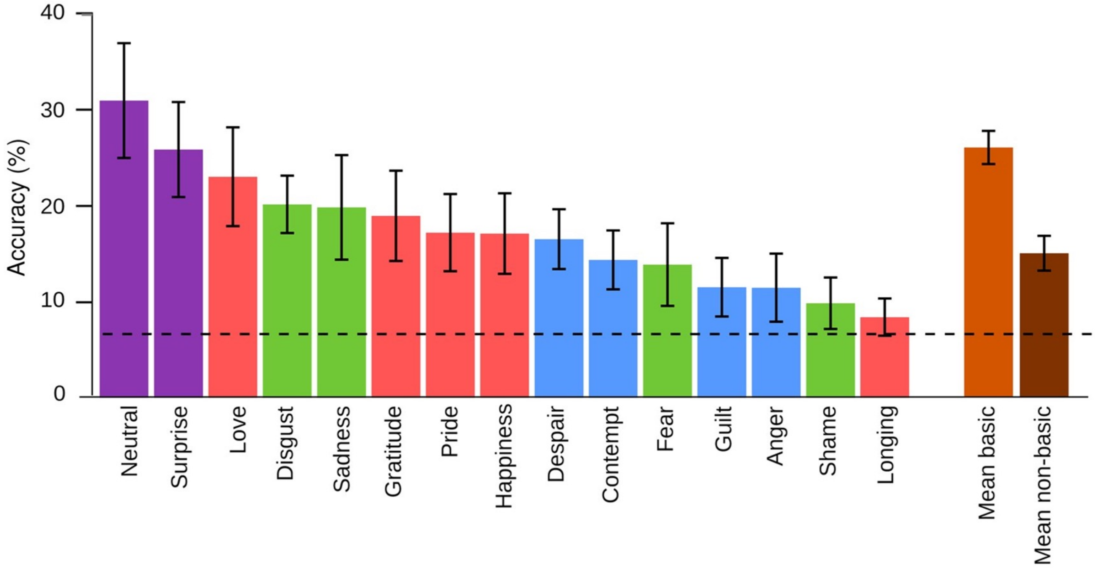
Step 6: Statistics



Accuracy
Category confusions
Important features

Classification accuracies

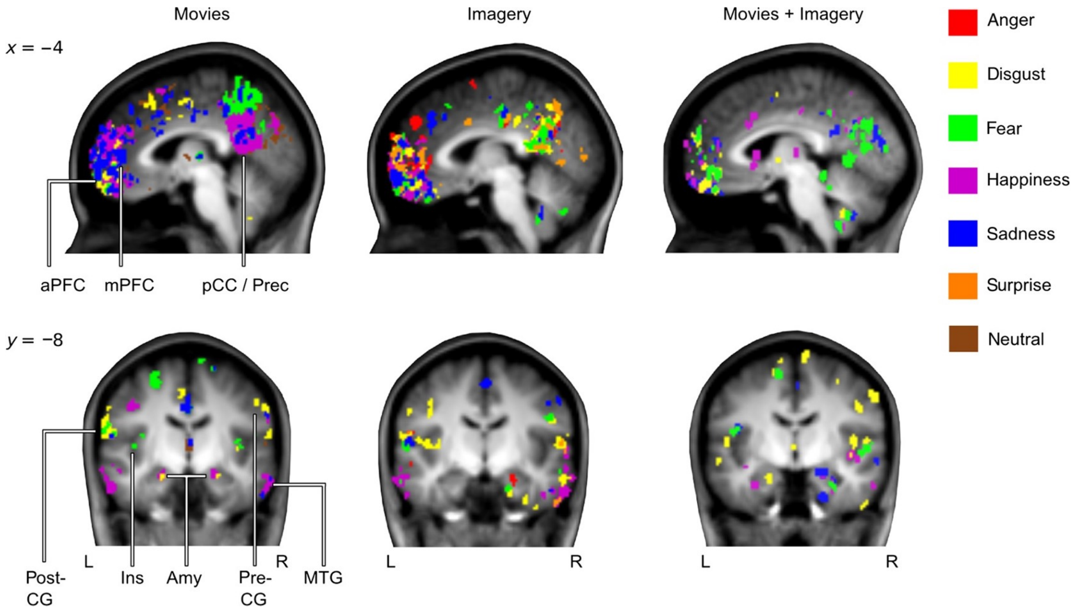
Accuracy: *significantly* above chance?
→ permutation tests



Saarimäki et al. (2018) *SCAN*

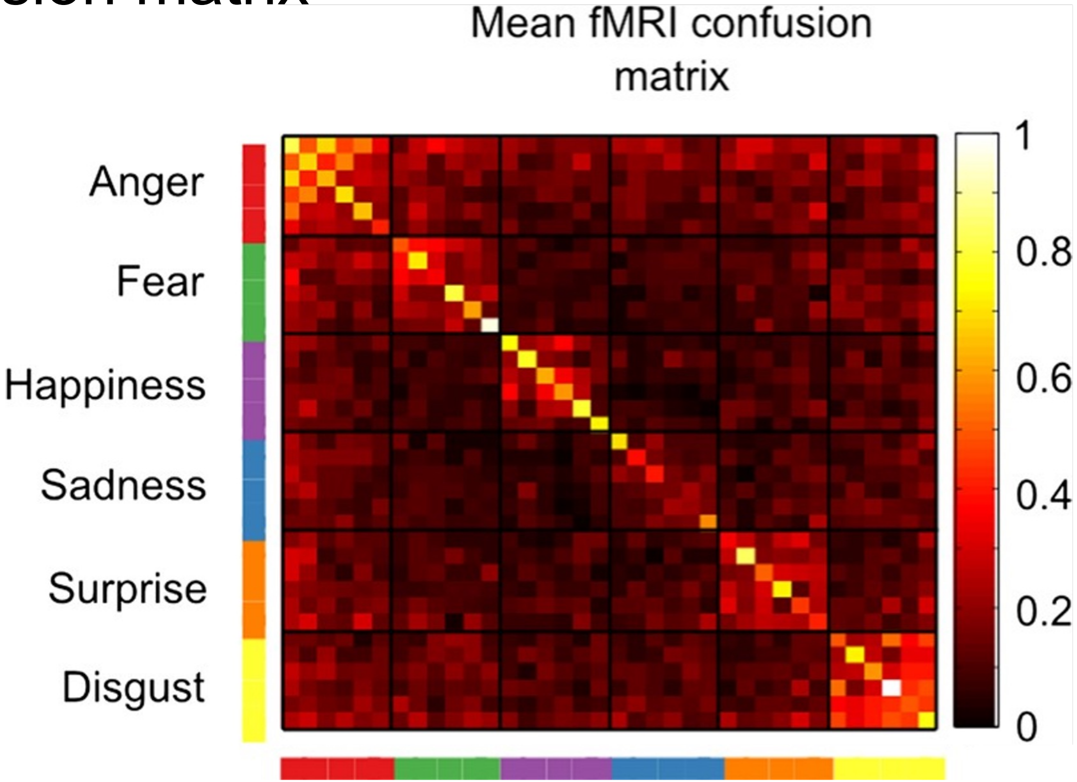
Importance maps

Features contributing to classification: importance maps



Saarimäki et al. (2016) *Cereb Cortex*

Confusion matrix



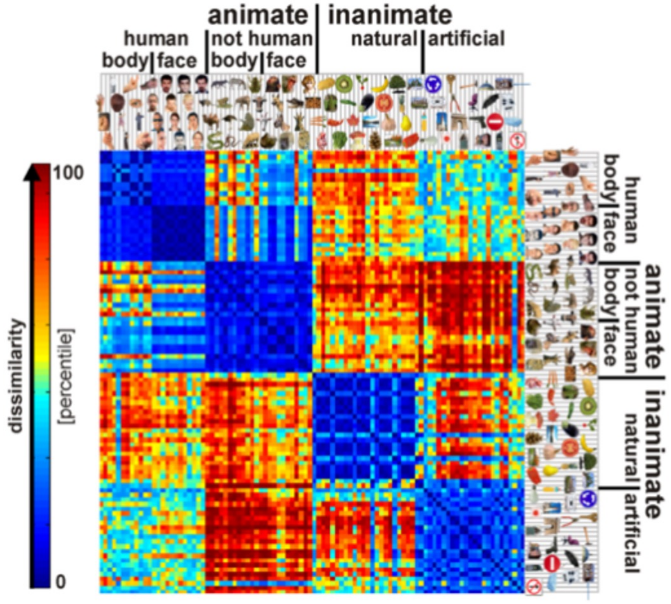
Saarimäki et al. (2016) *Cereb Cortex*

Representational similarity analysis (RSA)

RSA = A multivariate pattern analysis method to investigate the content and format of representations.

Pattern recognition:
are two patterns different?

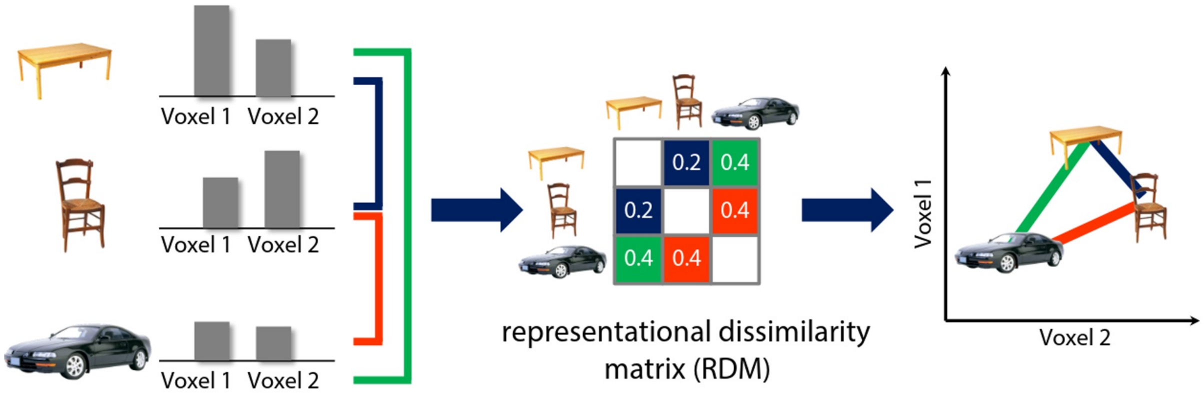
Representational similarity analysis:
how are the two patterns different?



Kriegeskorte et al. (2008) *Neuron*

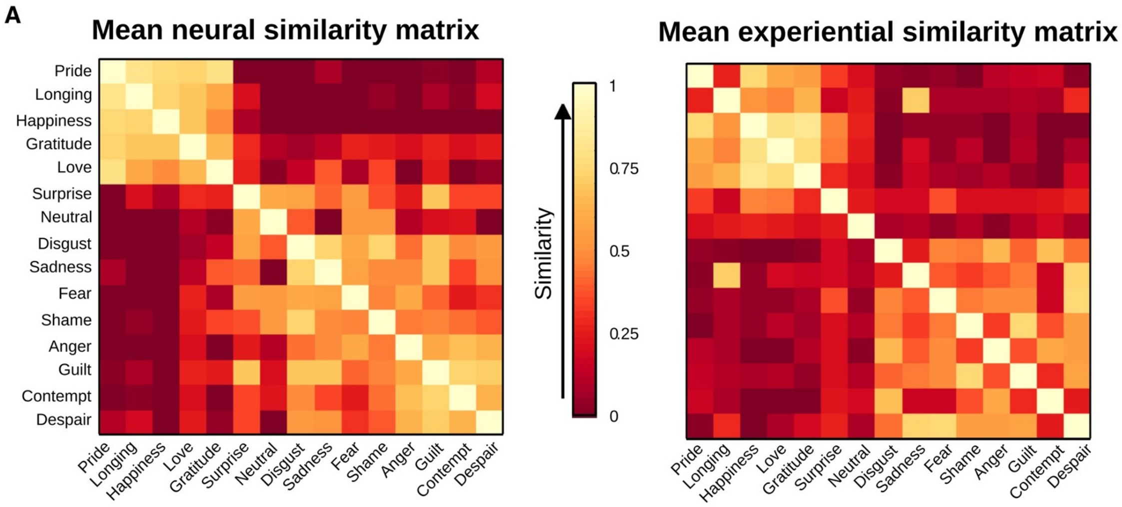
Representational similarity analysis

In RSA, we take our multivariate patterns (e.g. voxels) and calculate pairwise dissimilarities (e.g. Euclidean distance or 1 - Person's *r*)



Representational similarity analysis

Representational similarities can be used e.g. for testing models of cognition or for comparing different types of data.



Saarimäki et al. (2018) *SCAN*

More resources

Experimental design, methodological choices etc:
https://fmrif.nimh.nih.gov/public/other-courses/mvpa

Conceptual lectures by Rebecca Saxe:
https://cbmm.mit.edu/fmri-bootcamp

Toolboxes & tutorials e.g.:
https://brainiak.org/tutorials/
http://www.pymvpa.org/tutorial.html