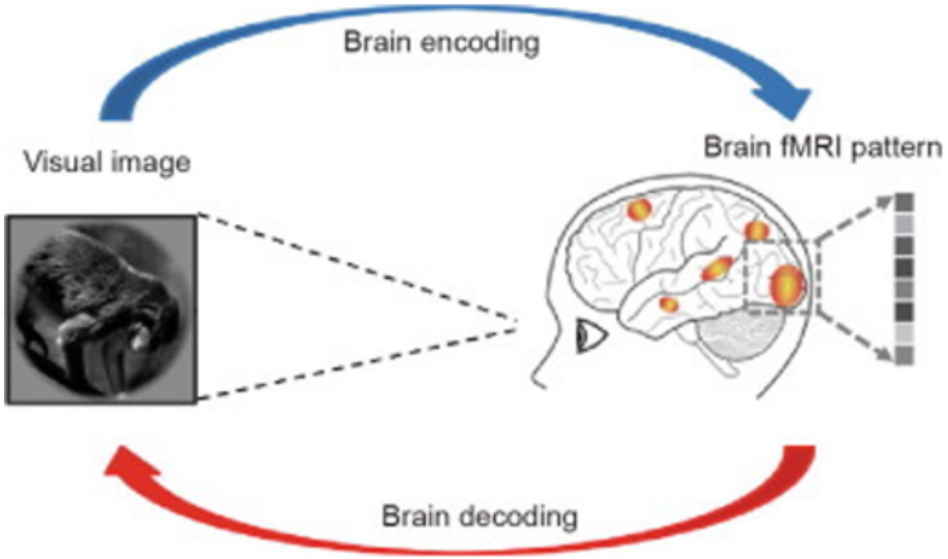
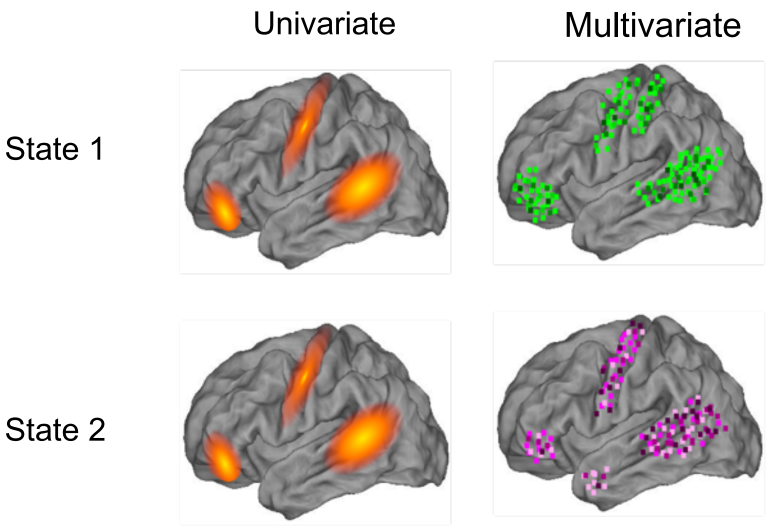
Statistical Pattern Recognition with fMRI

Heini Saarimäki
Postdoctoral Fellow
Human Information Processing Laboratory
Tampere University

Multivariate / multivoxel pattern analysis (MVPA), Pattern recognition, Machine learning, Decoding, Classification, ...





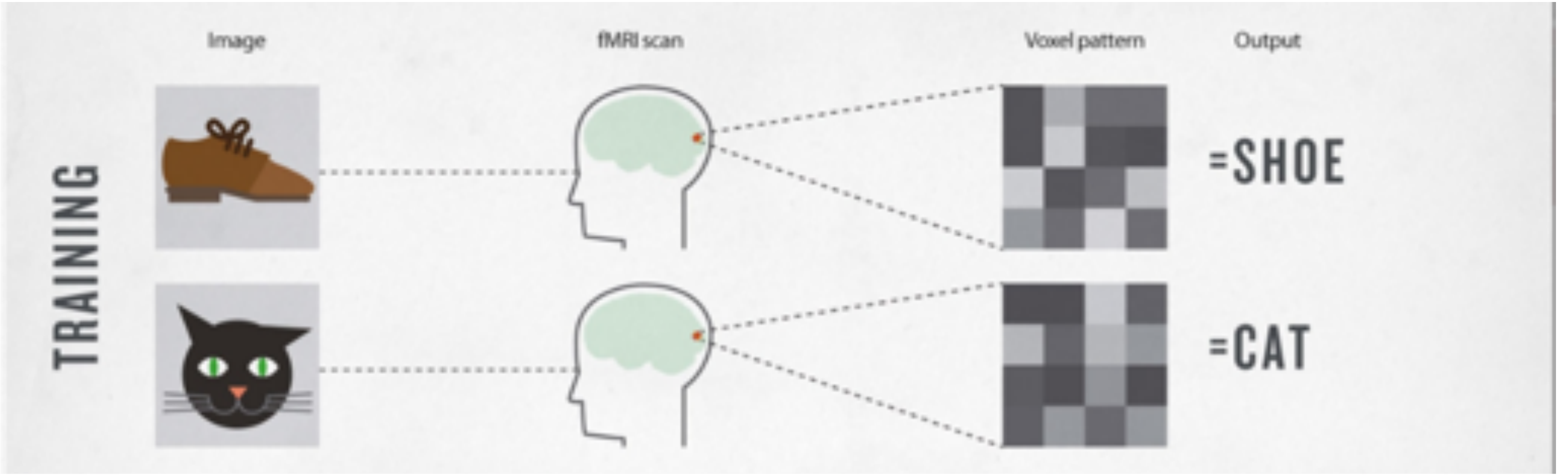
Outline

WHY: When and why to use classification methods

WHAT: Nuts and bolts of pattern classification

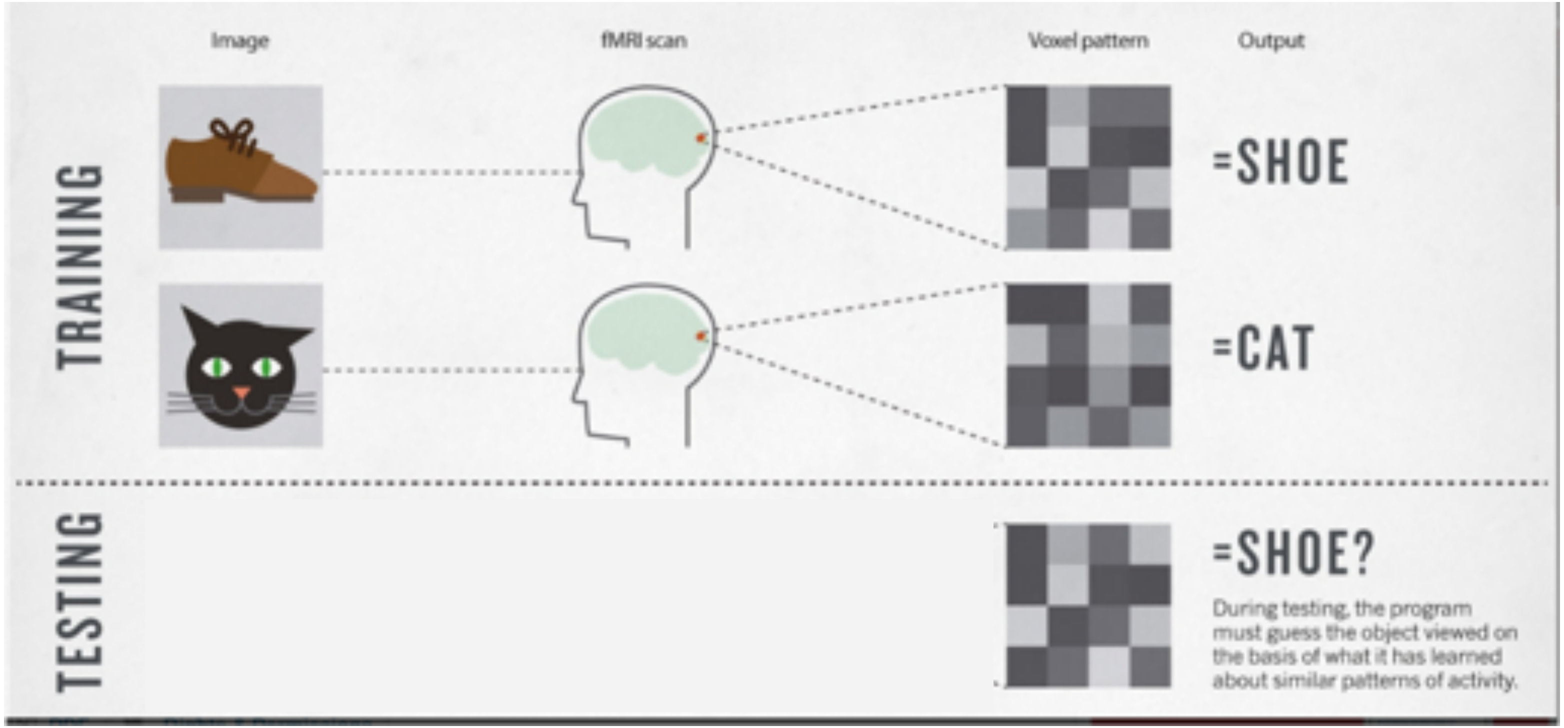
HOW: Overview of the workflow

Simplified framework:



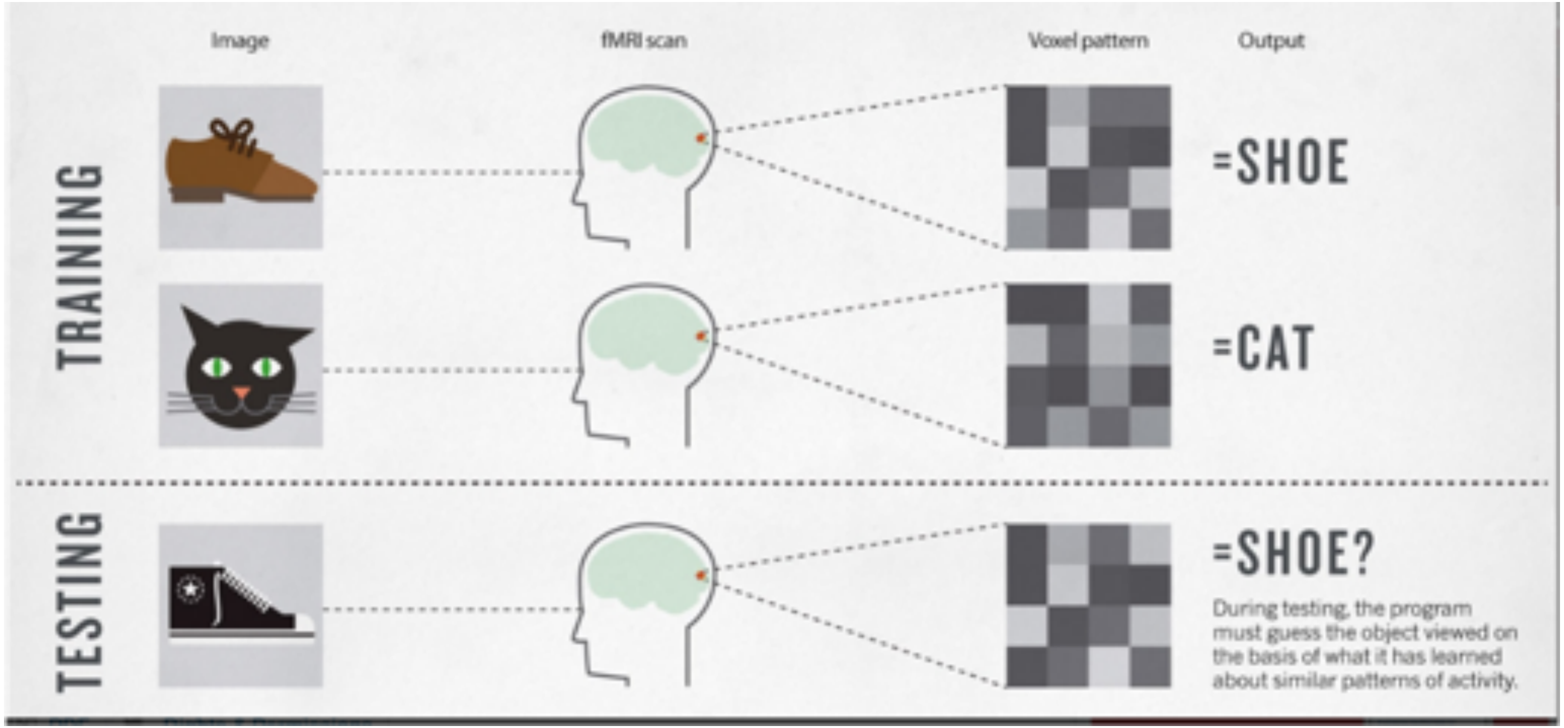
e.g. Haxby et al. (2001) *Science*
Figure credits: Smith (2013) *Nature*

Simplified framework:



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

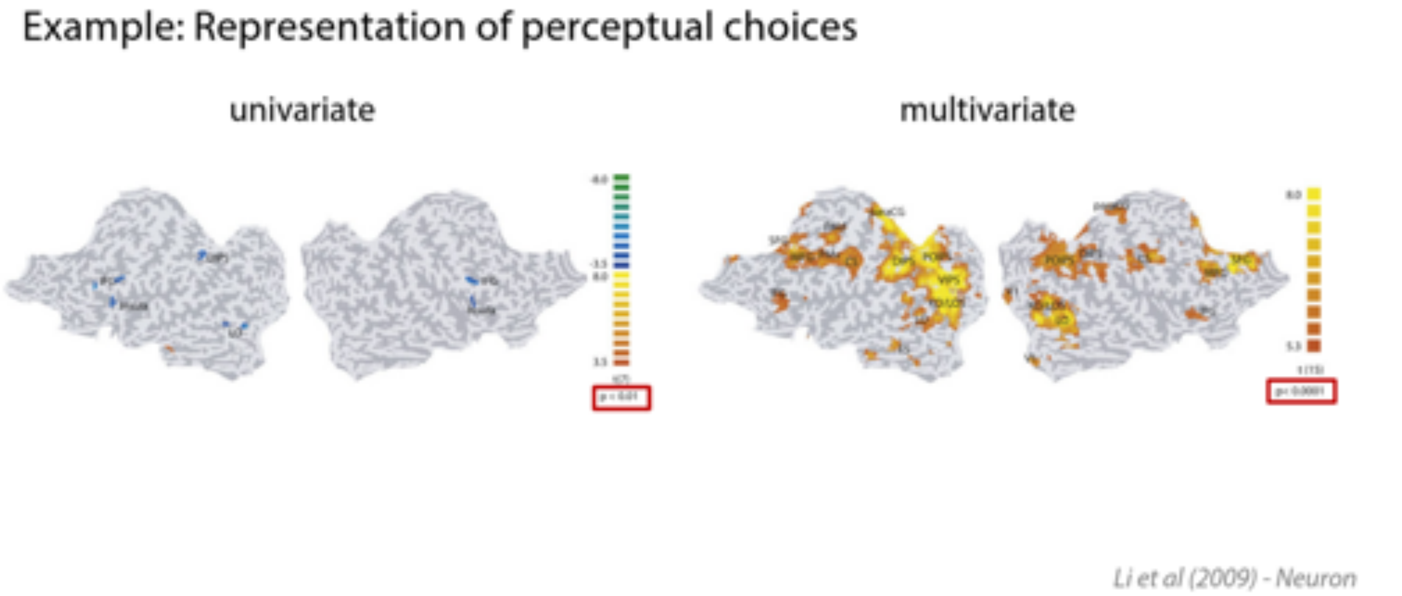
Simplified framework:



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

Why pattern recognition?

Higher sensitivity when compared to univariate analyses → detects pattern differences when activation overlaps

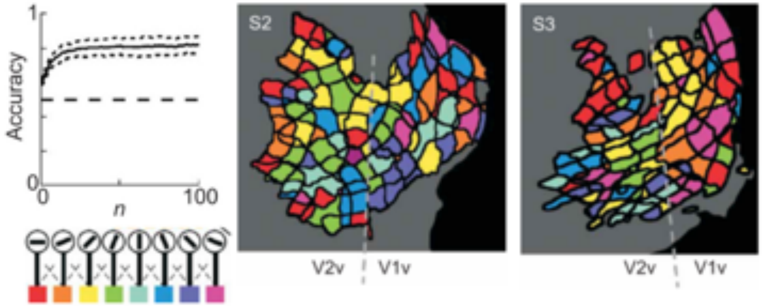


Why pattern recognition?

Higher sensitivity when compared to univariate analyses → pattern differences when activation overlaps

Can study representational content in a brain region rather than general activation (also activity vs information)

Example: Representation of orientations



Kamitani & Tong (2005) – Nat Neurosci, Haynes & Rees (2005) – Nat Neurosci

Why pattern recognition?

Higher sensitivity when compared to univariate analyses → pattern differences when activation overlaps

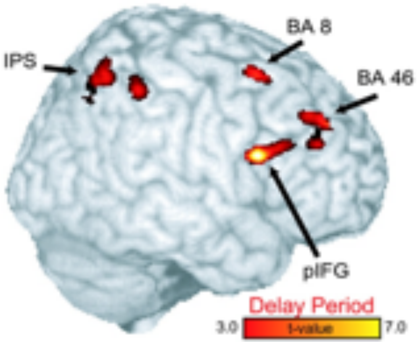
Can study representational content in a brain region rather than general activation (also activity vs information)

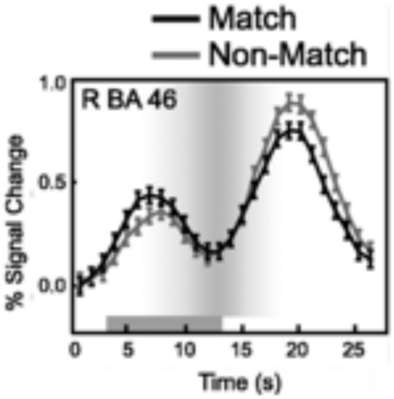
Methods readily available from machine learning

The principle is relatively easy to understand

Activity vs information

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

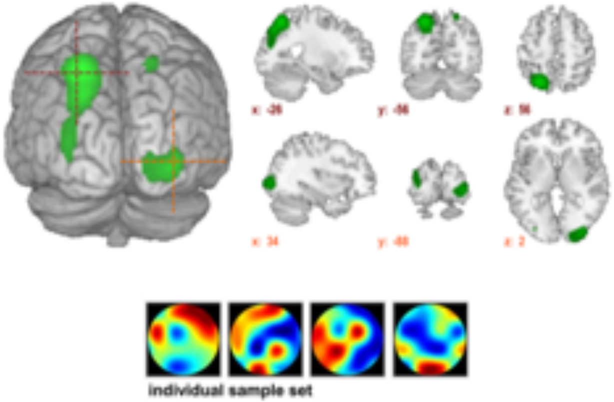


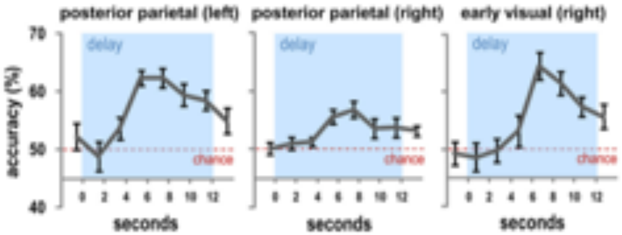


Curtis et al. (2004) – J Neurosci

Activity vs information

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





Christophel, Hebart, & Haynes (2012) – J Neurosci, but see Ester et al. (2015) – Neuron

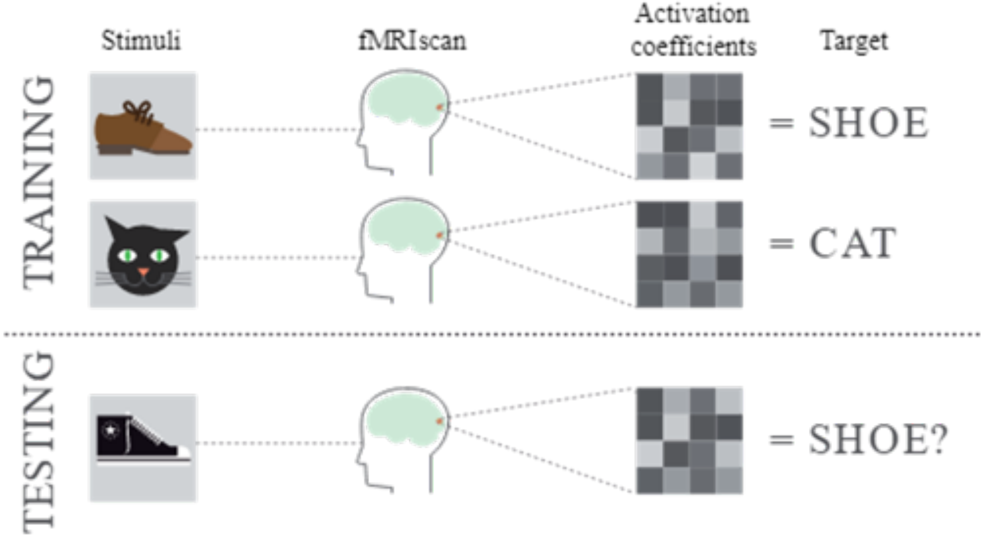
Applications

Task classification, testing theories of representational structure

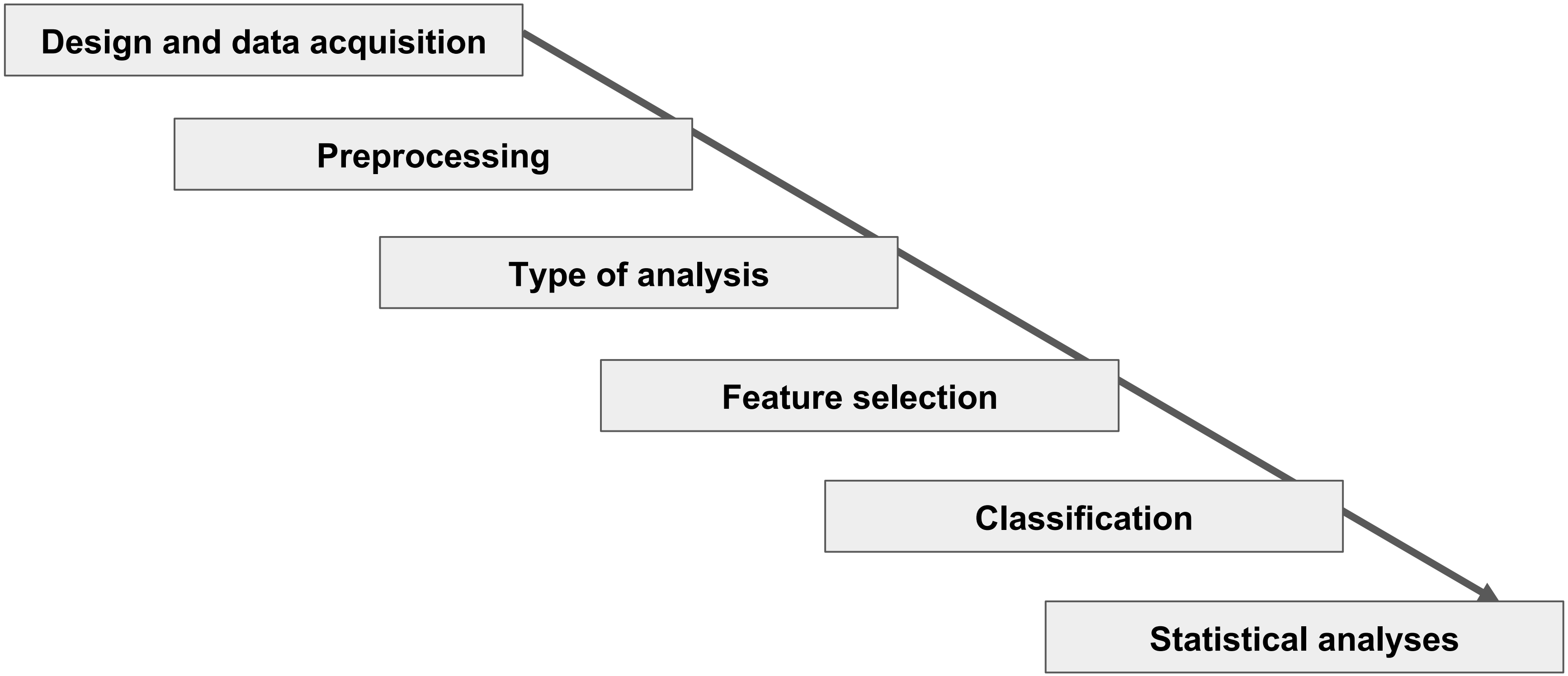
Medical diagnosis classification

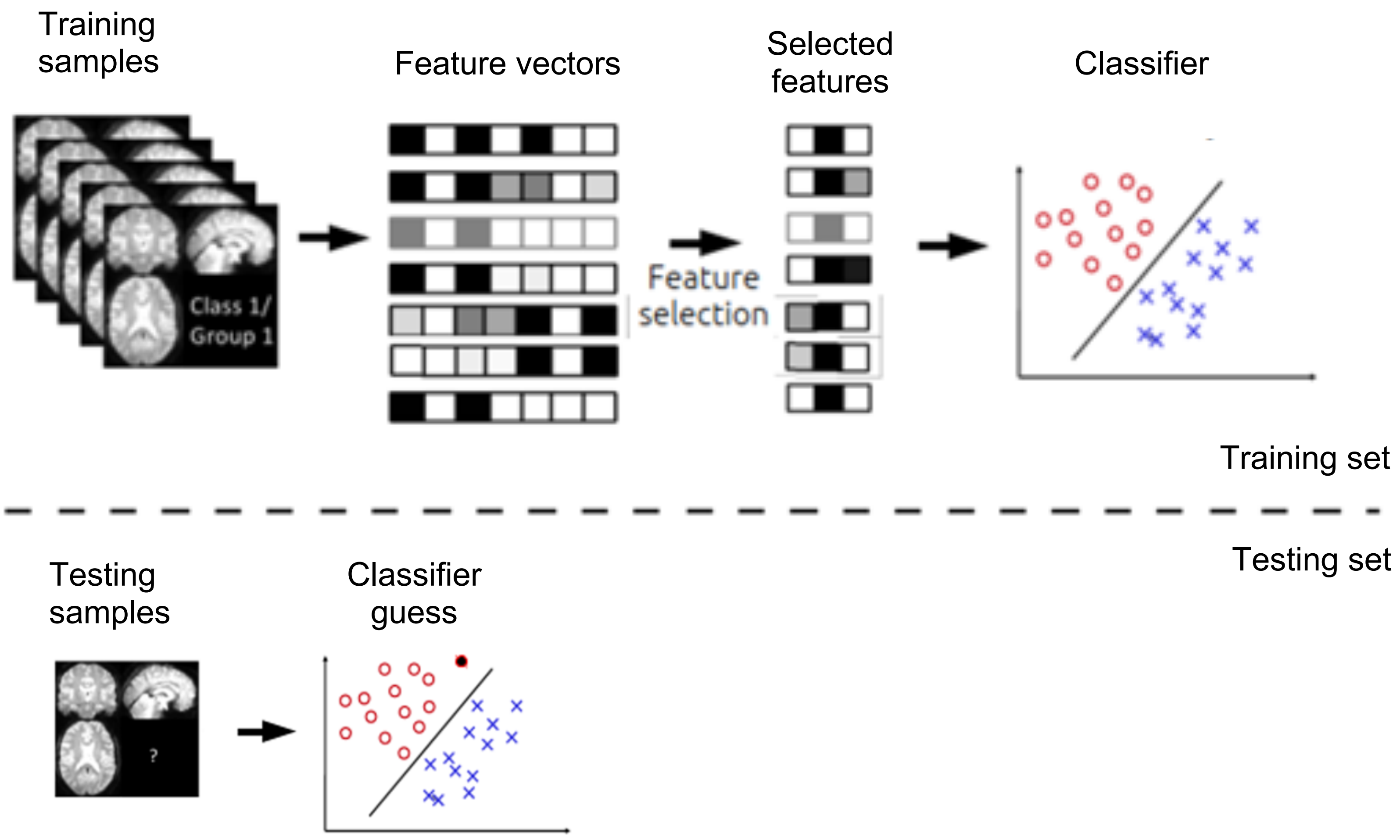
Medical Diagnosis

Brain-Computer-Interface



Overview of the workflow




Training
samples
Feature vectors
Selected
features
Classifier
Class 1/
Group 1
Feature
selection
Training set
Testing set
Testing
samples
Classifier
guess
?

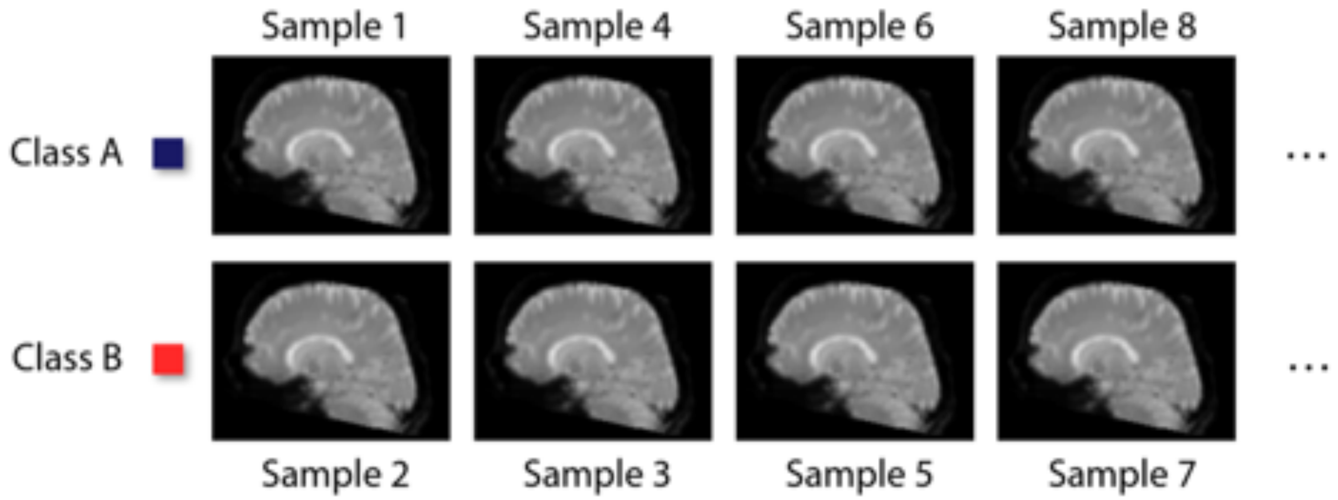
Step 1: Design and data acquisition

Categories

Key concepts: Sample

Samples are data that belong to a class

Examples: EPI volumes, beta volumes, VBM maps, EEG data

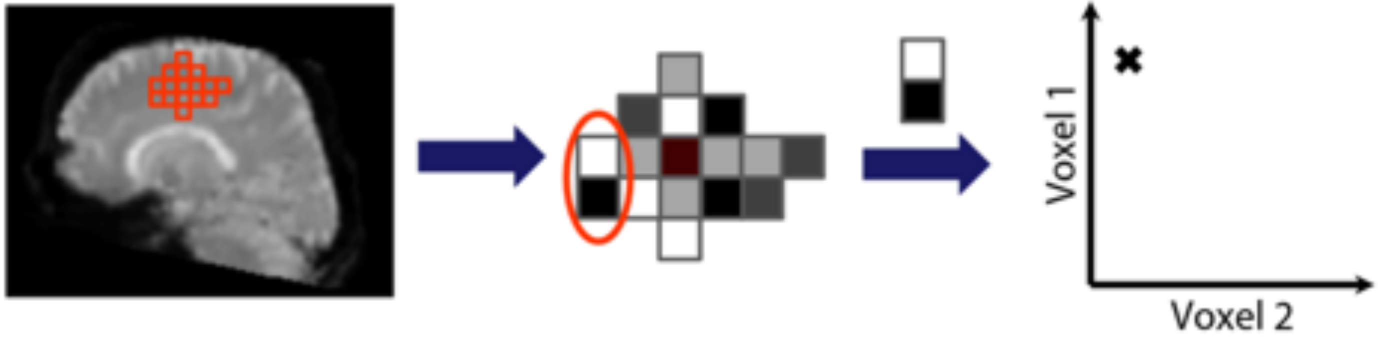


Key concepts: Feature

Each feature is a measured variables that can be used for classification

- Each feature (hopefully) aids the classification process, by contributing signal and/or suppressing noise
- Each feature spans up a dimension → they build the feature space

Examples: A voxel, connectivity graph, EEG channel



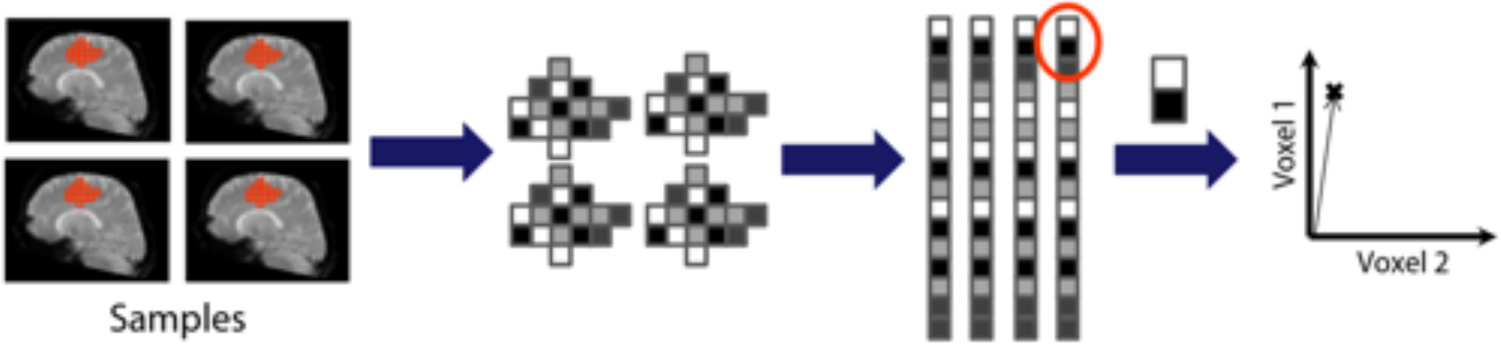
Key concepts: Pattern

A pattern is a sample for a set of features

A pattern is a point (or vector) in p-dimensional space (p is # of features)

Alternative uses of term "pattern" with different meaning:

- Prototypical pattern (i.e. the true class mean)
- Discriminating pattern (function that discriminates classes)



Key concepts: Label

A label denotes the class membership of a pattern with a number

For classification the number is categorical and often arbitrary (some classifiers require 0 and 1 or -1 and 1)

For regression the number denotes a continuous number which is the regression target

Class A　　Class B

Label: 1　　Label: -1

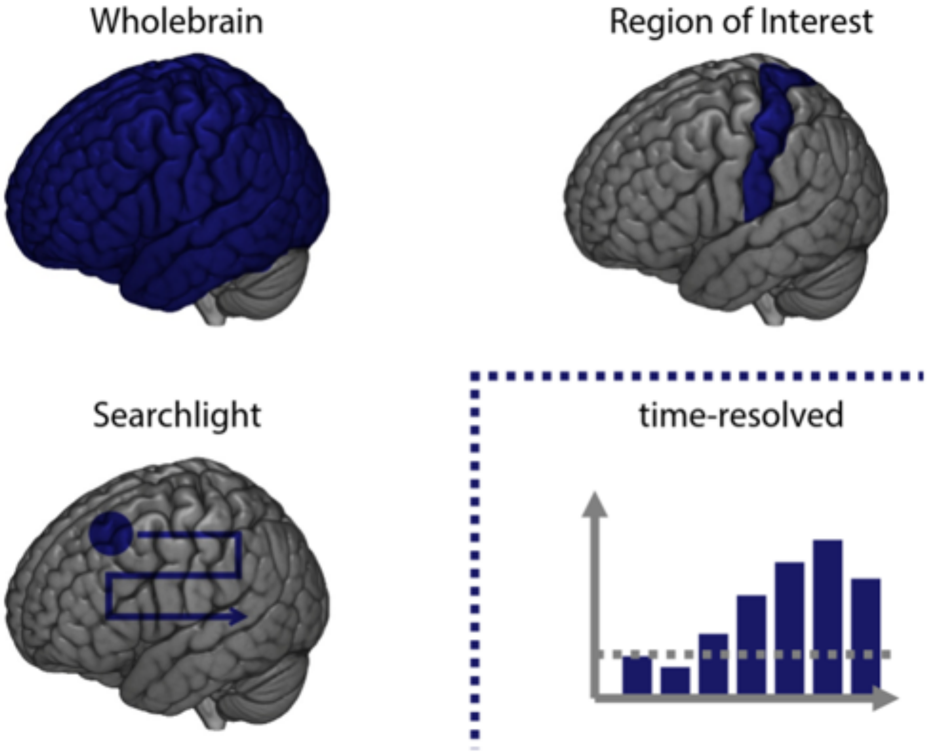
Step 2: Preprocessing

Minimal preprocessing

No smoothing

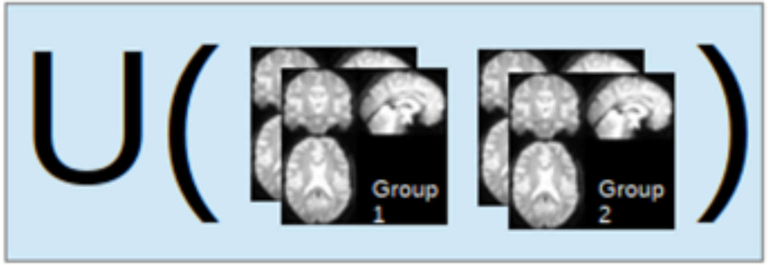
Normalization of signal intensities

Step 3: Type of analysis

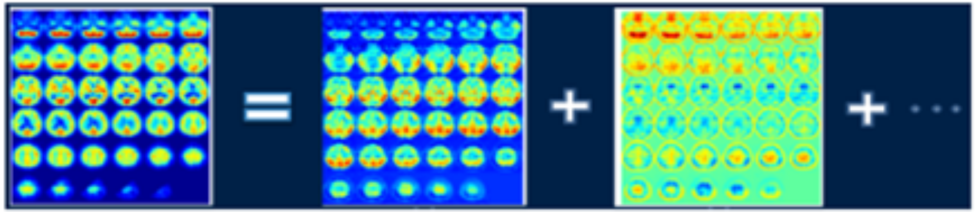


Step 4: Feature selection

ANOVA

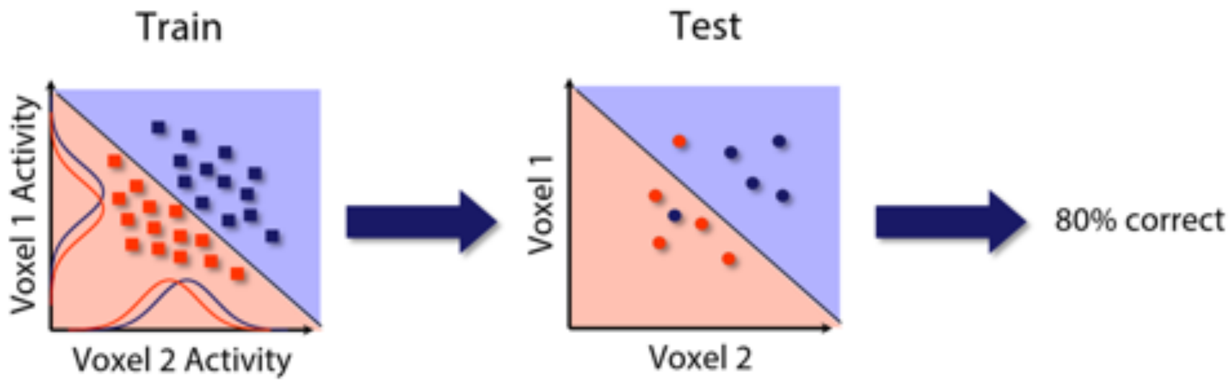


PCA



Step 5: Classification

1. Separate data in training and test data
2. "Train" a classifier (i.e. find a good separating line)
3. Apply this classifier (i.e. the line) to test data
4. Result: Accuracy of test data (i.e. % correct prediction)



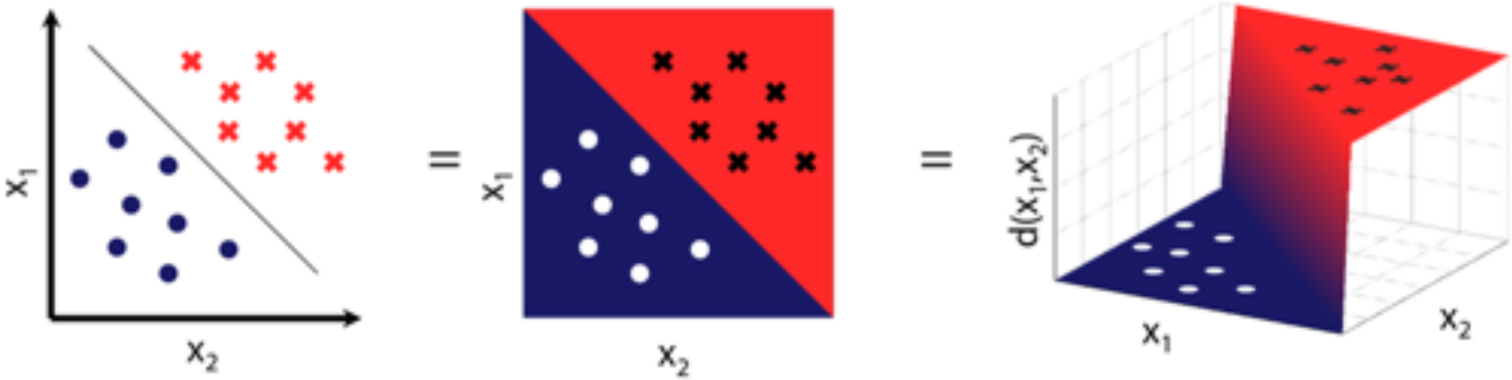
Key concepts: Classifier

A function that separates feature space

Example for one sample with two features: $f(x_1, x_2) = -0.5$

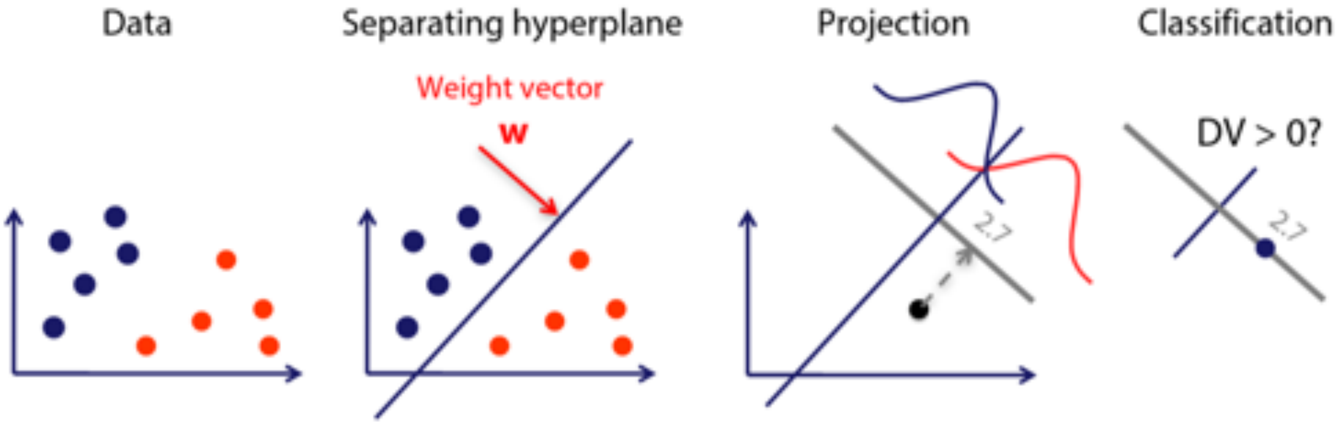
This decision value f is then binarized in a decision function:

$if\ f(x_1, x_2) > 0:\ \boldsymbol{d(x_1, x_2) = 1};\ if\ f(x_1, x_2) \leq 0:\ \boldsymbol{d(x_1, x_2) = -1}$



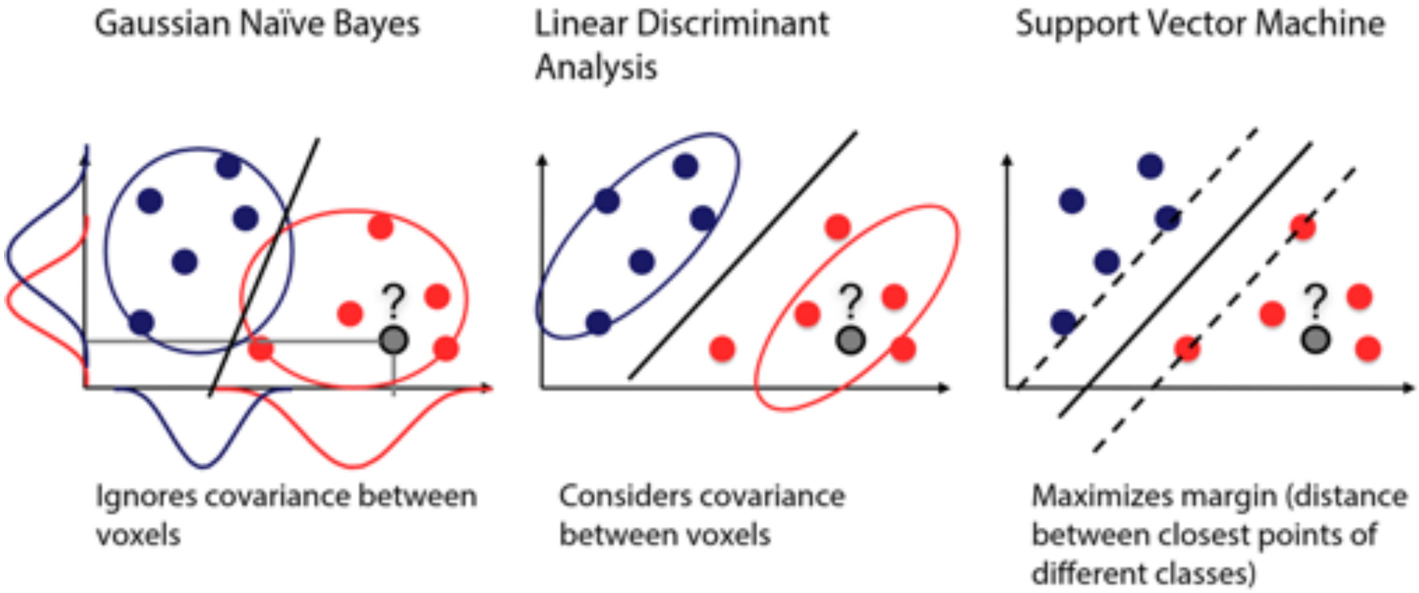
Step 5: Classification

Geometric intuition



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

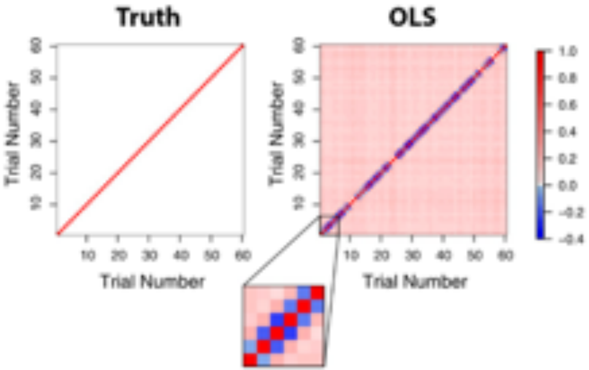
Step 5: Classification - linear classifiers

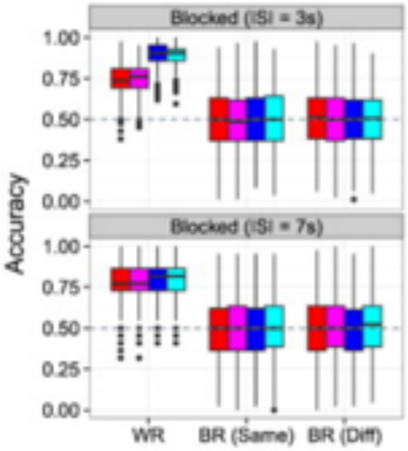


Step 5: Classification - cross-validation

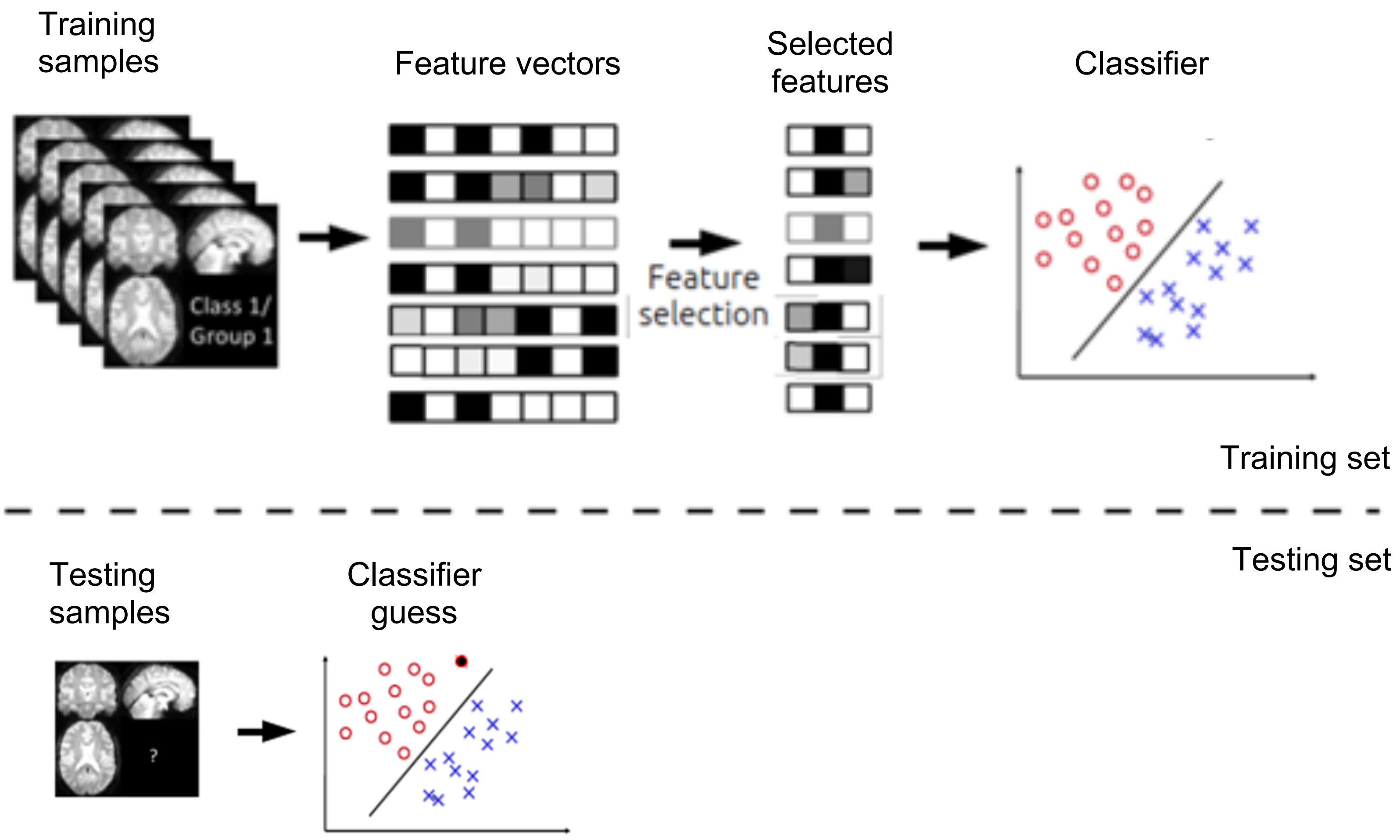
Typical Analysis: Leave-one-run-out cross-validation

Reason: Non-independence within run can bias results

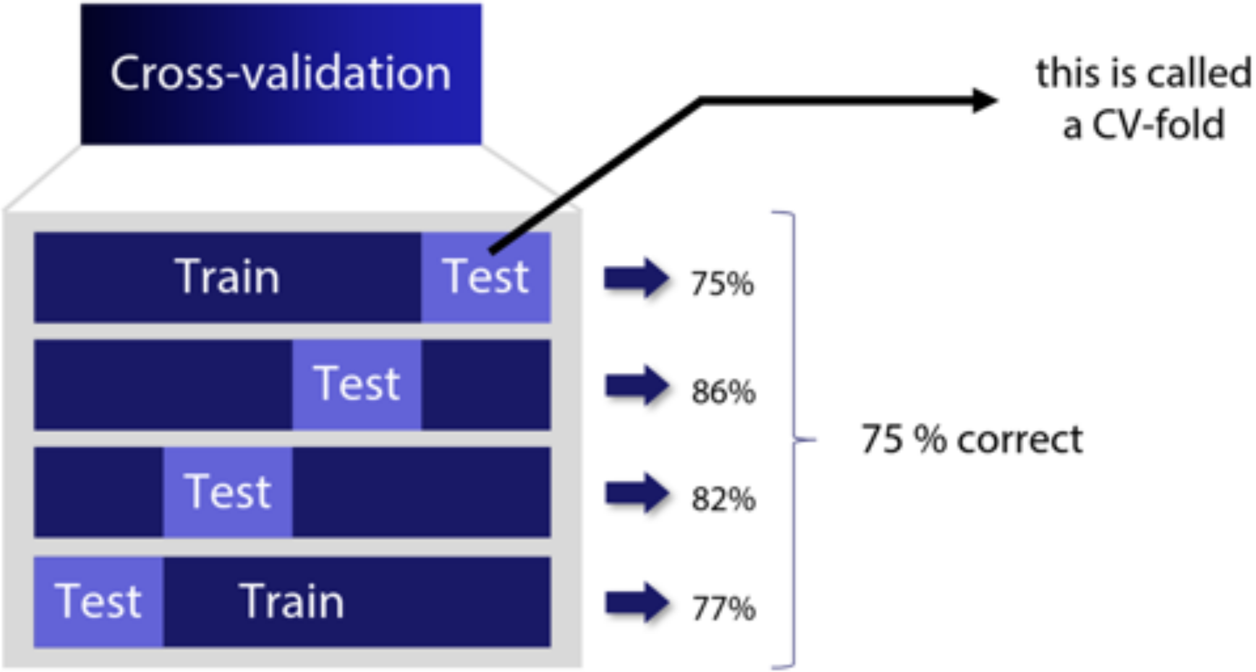




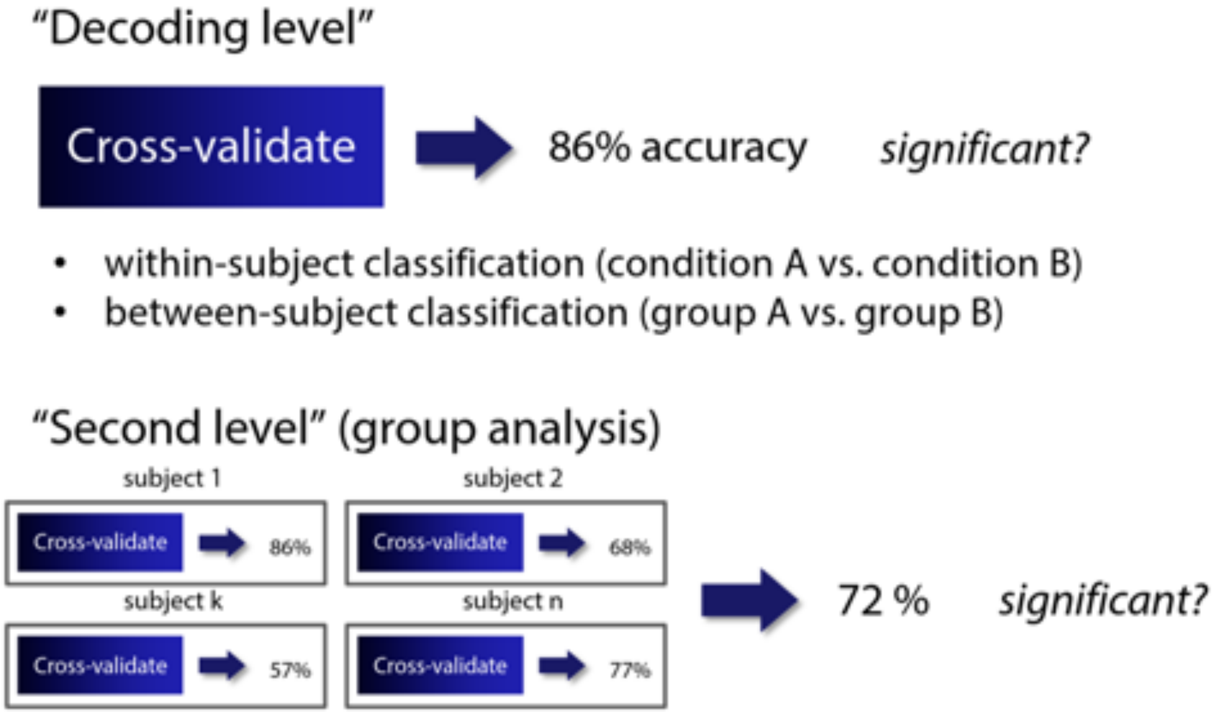
Mumford et al (2014) – Neuroimage


Training samples
Feature vectors
Selected features
Classifier
Class 1/ Group 1
Feature selection
Training set
Testing set
Testing samples
Classifier guess
?

Step 5: Classification - cross-validation



Step 6: Statistics - classifier accuracy



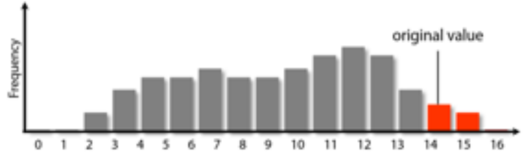
Step 6: Statistics - permutation tests

Procedure:

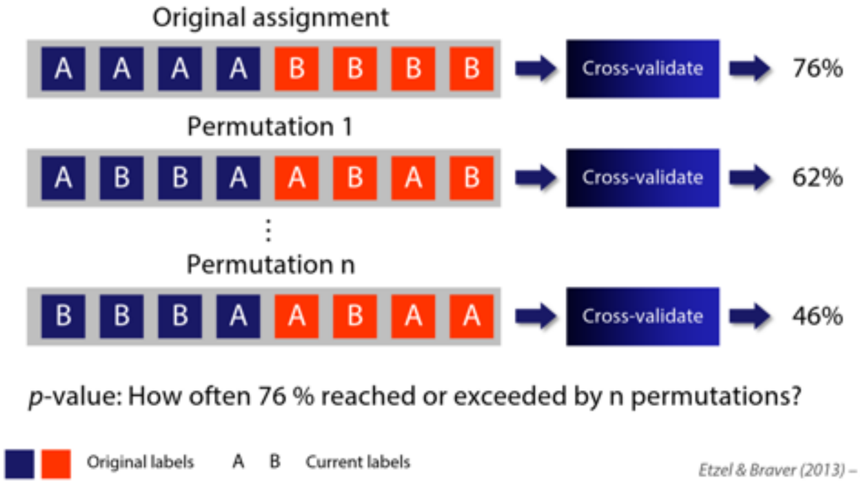
1. Calculate statistic (e.g. accuracy) using normal procedure
2. Permute labels and repeat same procedure
3. Repeat n times or until exhaustion

For exhaustive test: *p*-value is $\frac{k}{n}$ where *k* is number of permutations with equal or higher accuracy and *n* is number of all permutations (includes original result)

For non-exhaustive test (aka Monte-Carlo permutation test): $\frac{k+1}{n+1}$



Typical procedure: Label permutation (i.e. random shuffling)

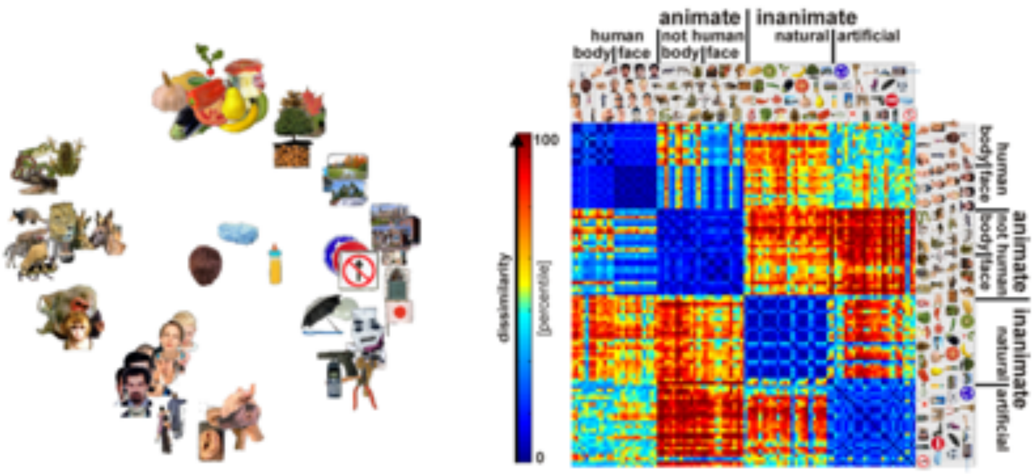


Etzel & Braver (2013) – PRNI

Step 6: Statistics - confusion matrix

Step 6: Statistics - representational similarity analysis

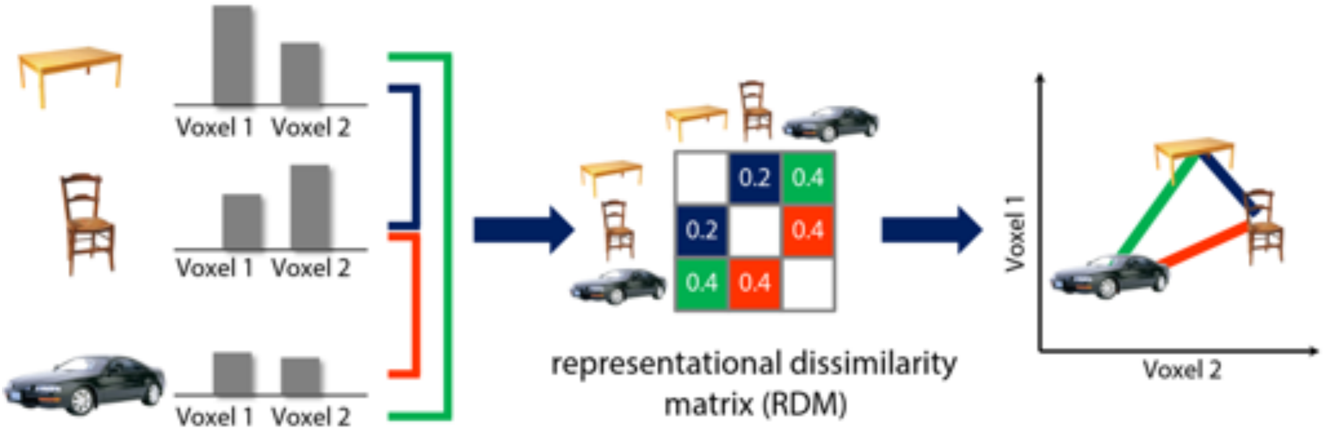
A multivariate pattern analysis method to investigate the *content* and *format* of representations



Kriegeskorte et al (2008) – Neuron; Kriegeskorte & Kievit (2013) – TICS

Step 6: Statistics - representational similarity analysis

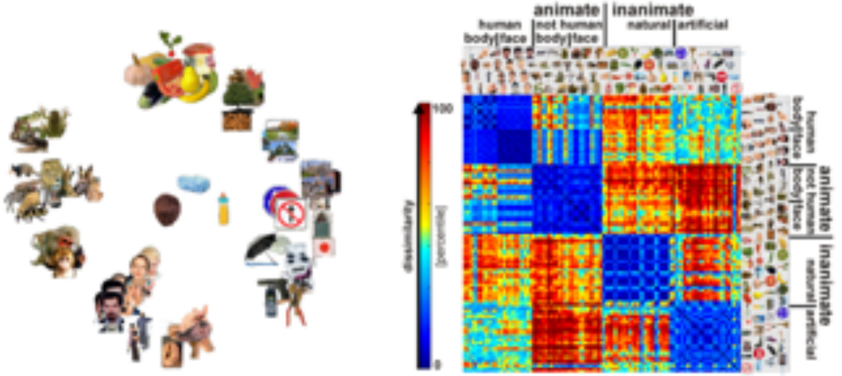
- In RSA, we take our multivariate patterns (e.g. voxels) and calculate pairwise dissimilarities (e.g. Euclidean distance or 1 – Pearson's *r*)



Step 6: Statistics - representational similarity analysis

1. Simple exploratory approach to characterize **multidimensional representations**

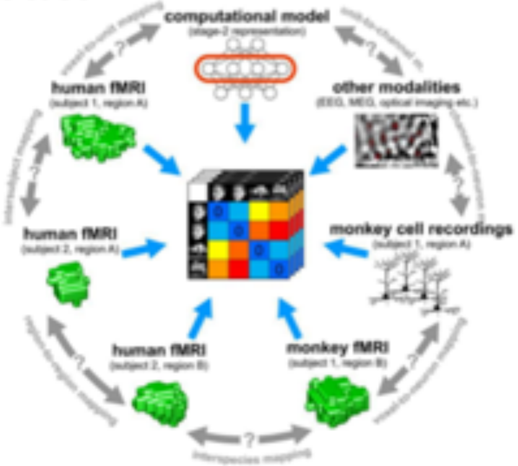
Example: How does human IT represent object categories?



Kriegeskorte et al (2008) – Neuron; Kriegeskorte & Kievit (2013) – TICS

Step 6: Statistics - representational similarity analysis

2. Representational (dis)similarity matrices can be seen as a **common language to study representations** across methods (MEG, fMRI, cell recording, …), brain regions, humans and species

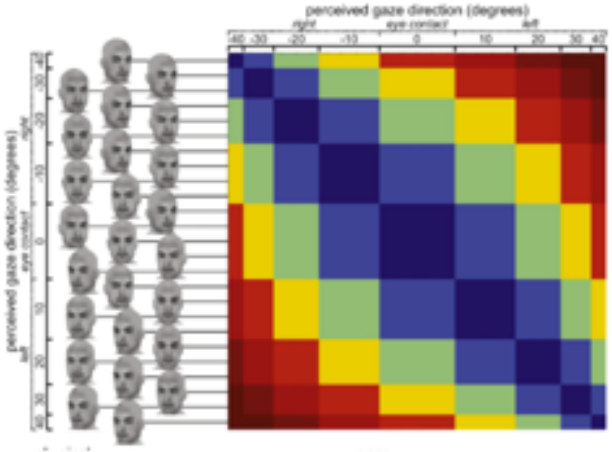


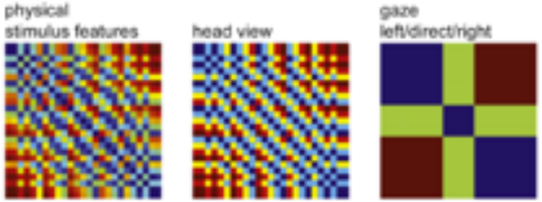
Kriegeskorte et al (2008) – Frontiers; Kriegeskorte & Kievit (2013) – TICS

Step 6: Statistics - representational similarity analysis

3. Representational similarities can be used for **testing models of cognition**

Example: Which facial features does a brain region represent?





Carlin et al (2011) – Curr Biol

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

Tutorials:
https://brainiak.org/tutorials/