fMRI classification analysis: a conceptual introduction

Marieke Mur

CBU, march 2014

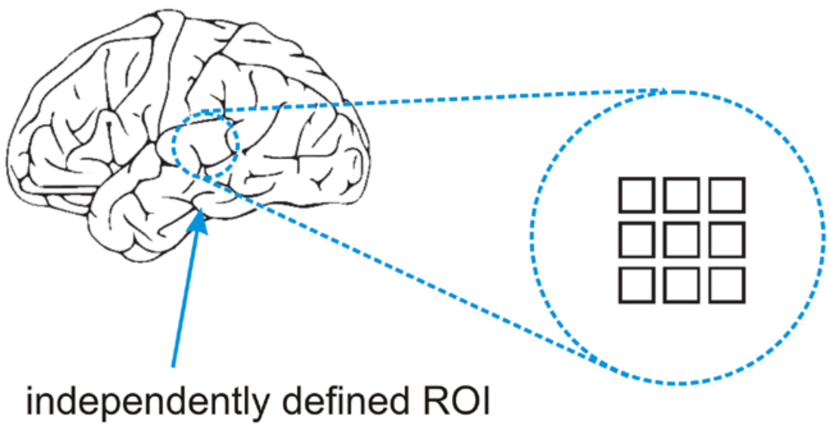
Overview

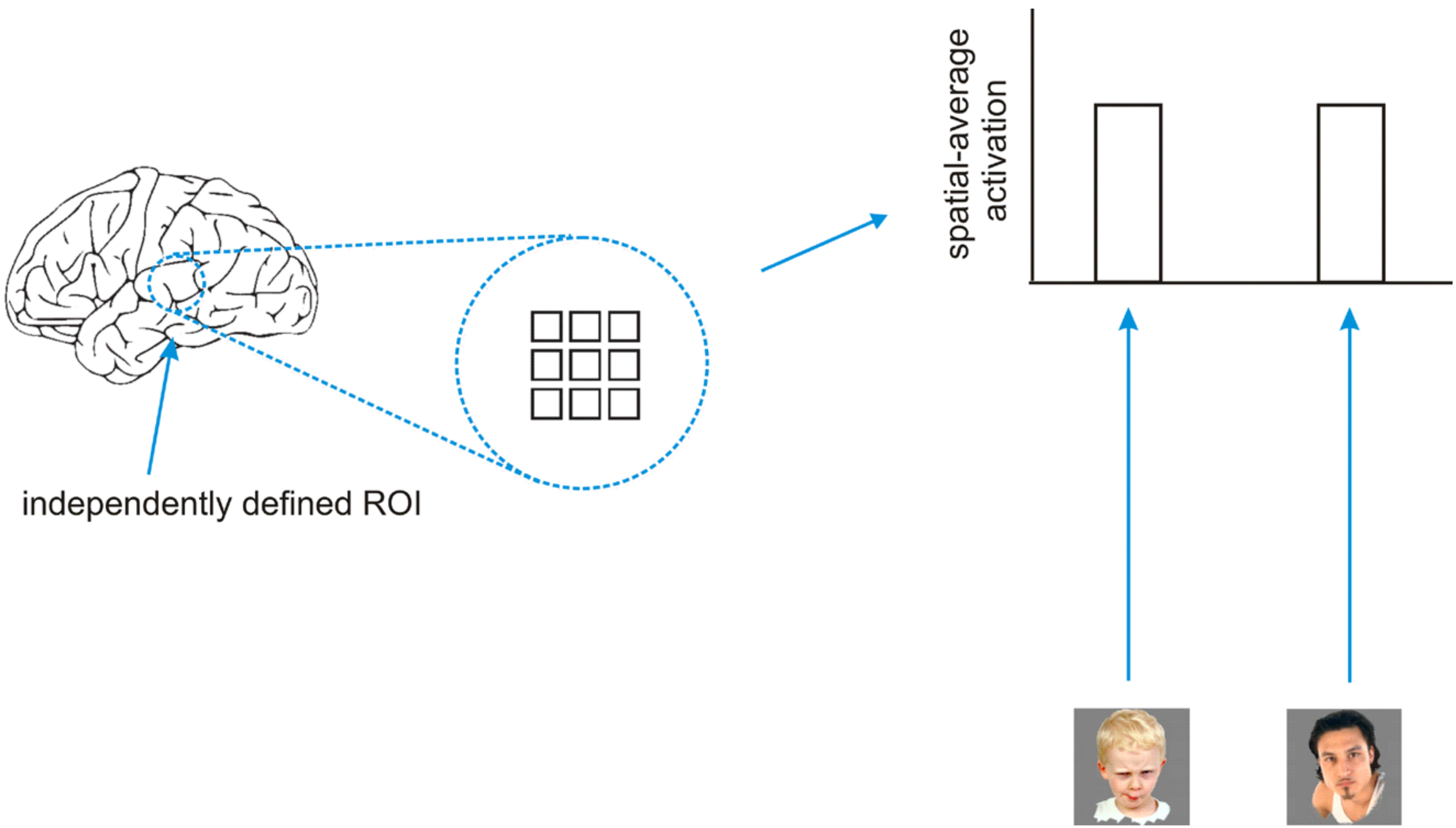
- Why classification analysis?
- Linear classification: the basic idea
- Linear classification: different classifiers
- Do it yourself: six steps
  - step 1: split data and preprocess
  - step 2: estimate single-subject activity patterns
  - step 3: select voxels
  - step 4: train the classifier
  - step 5: test the classifier
  - step 6: statistical inference
- Toolboxes
- Literature

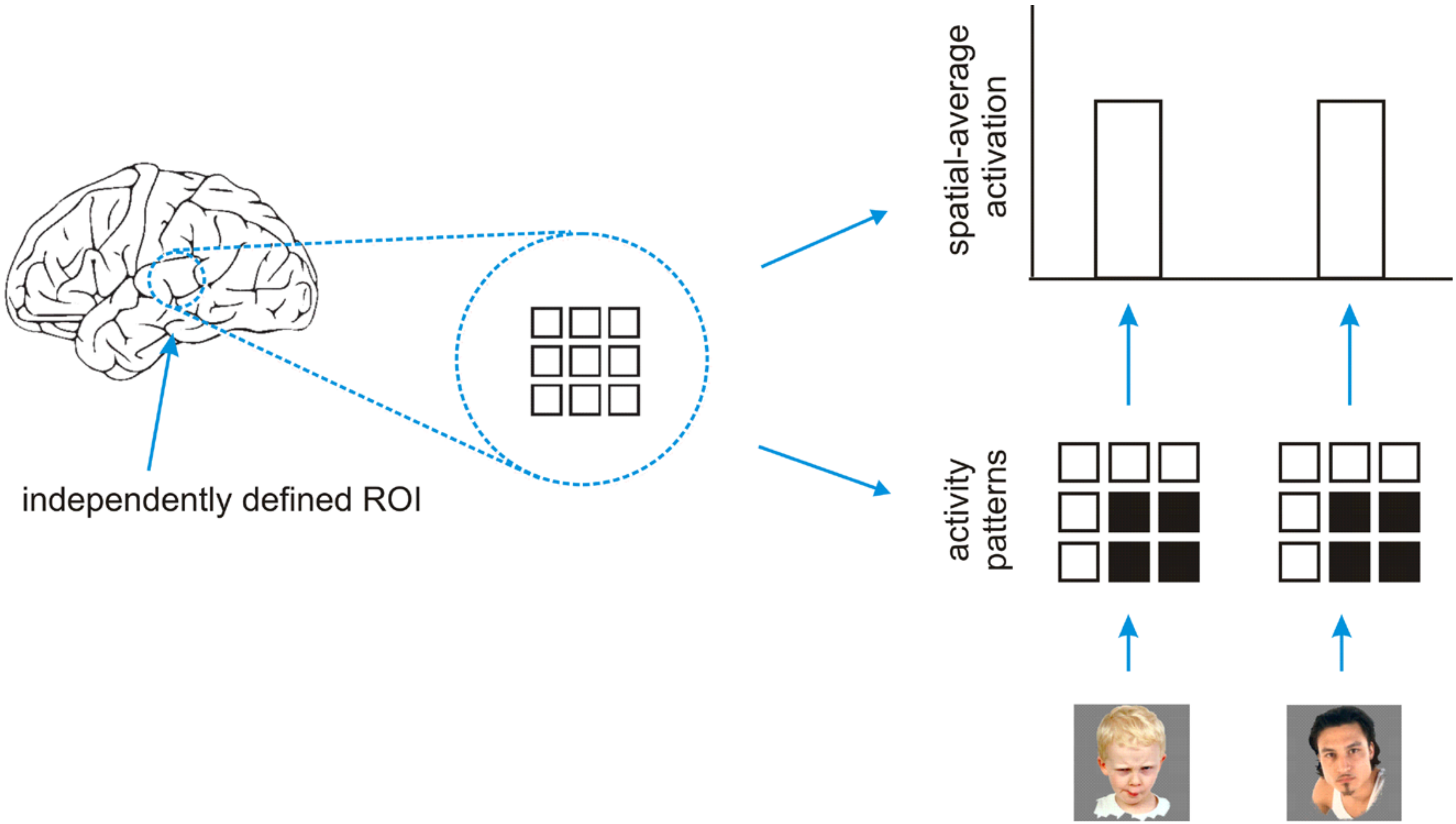
Overview

- **Why classification analysis?**
- Linear classification: the basic idea
- Linear classification: different classifiers
- Do it yourself: six steps
    - step 1: split data and preprocess
    - step 2: estimate single-subject activity patterns
    - step 3: select voxels
    - step 4: train the classifier
    - step 5: test the classifier
    - step 6: statistical inference
- Toolboxes
- Literature

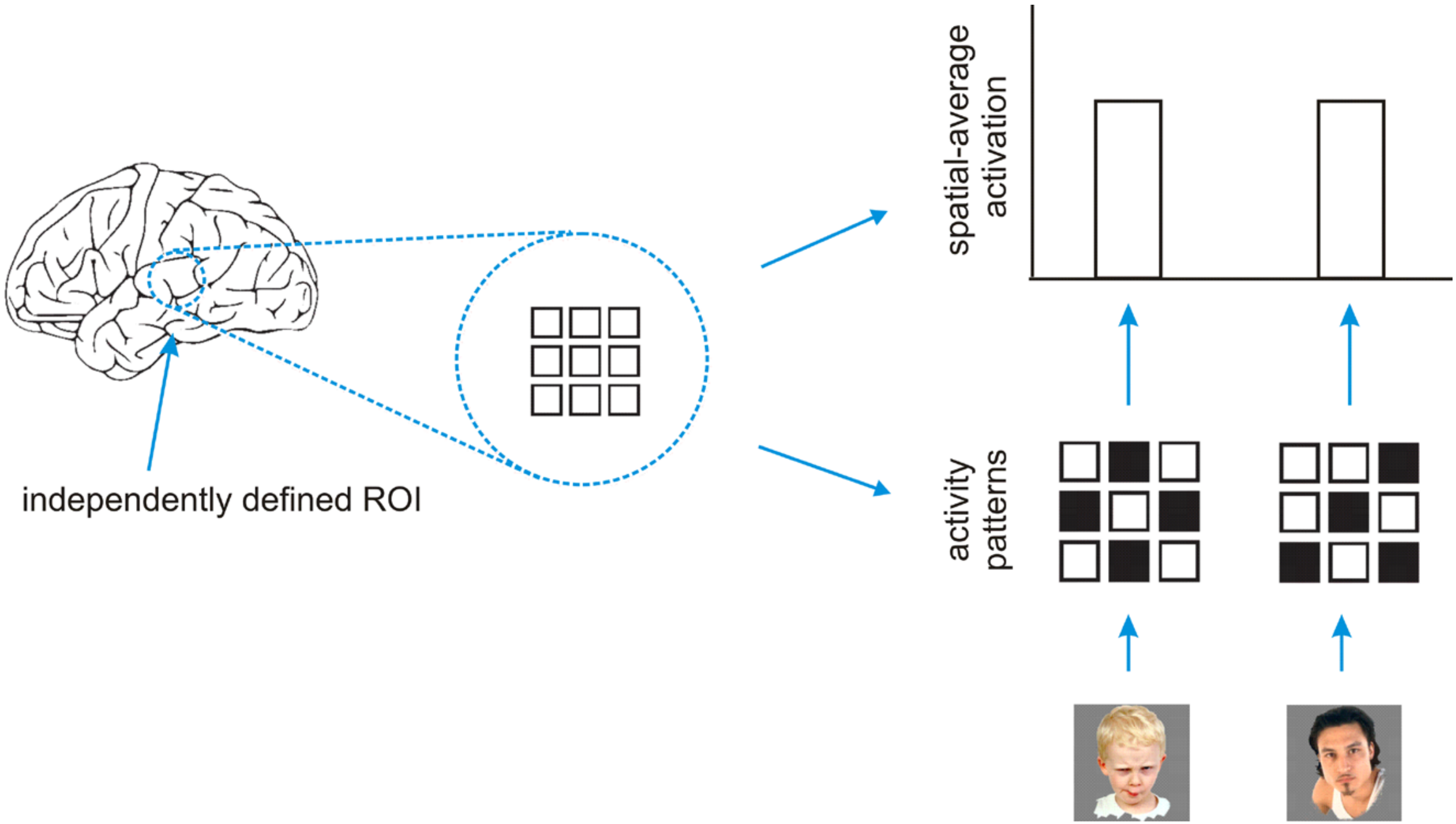
Activation-based analysis




spatial-average activation
independently defined ROI


independently defined ROI
spatial-average activation
activity patterns

Pattern-information analysis



Pattern-information analysis

Goal

Determine whether activity patterns elicited by different conditions are statistically discriminable.

How?

Multivariate analysis of variance (MANOVA)?

Pattern-information analysis

Goal

Determine whether activity patterns elicited by different conditions are statistically discriminable.

How?

Multivariate analysis of variance (MANOVA)?

Pattern-information analysis

Goal

Determine whether activity patterns elicited by different conditions are statistically discriminable.

How?

Approach pattern analysis as a classification problem.

Pattern classification

IF

we can classify the experimental conditions on the basis of the activity patterns better than chance

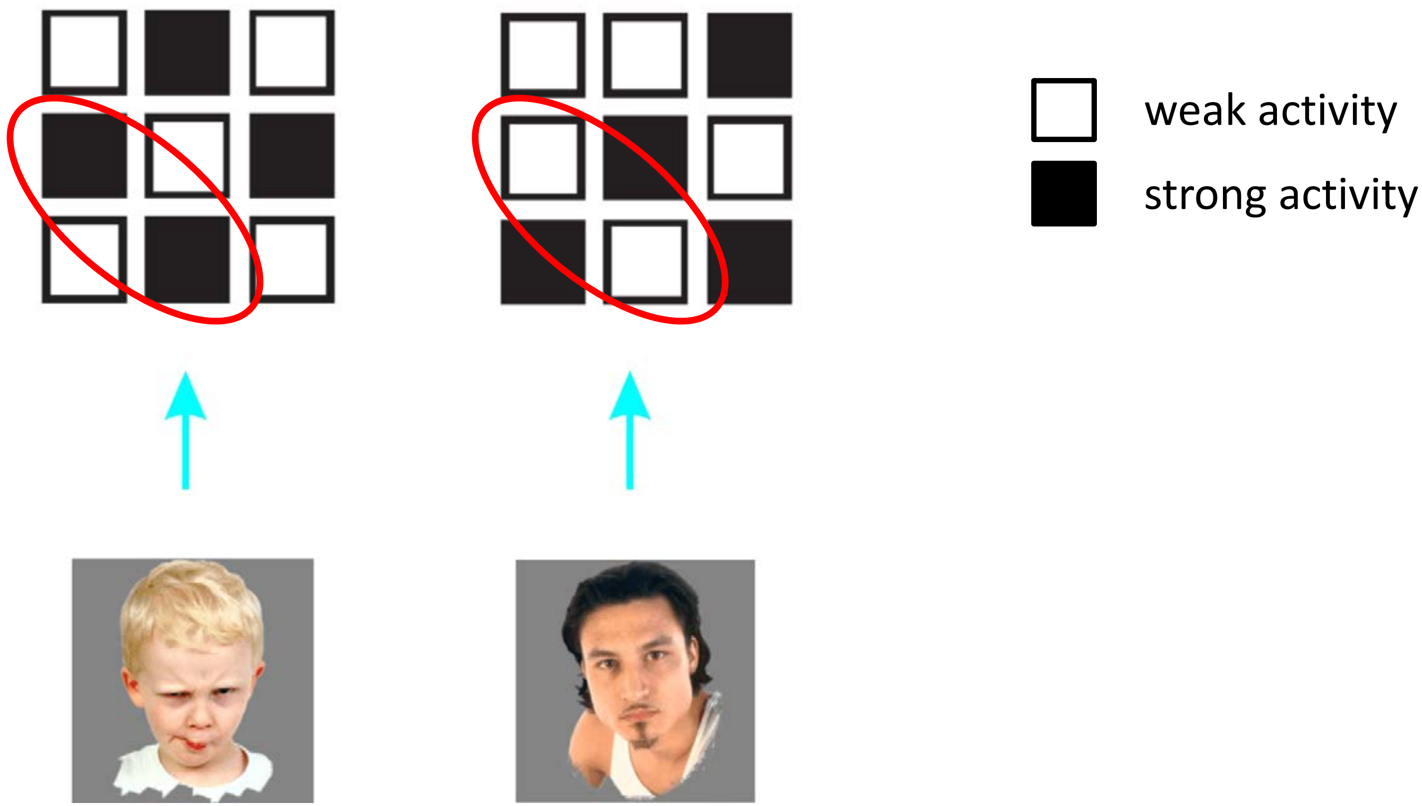
THEN

this indicates that the activity pattern carries information about the experimental conditions.

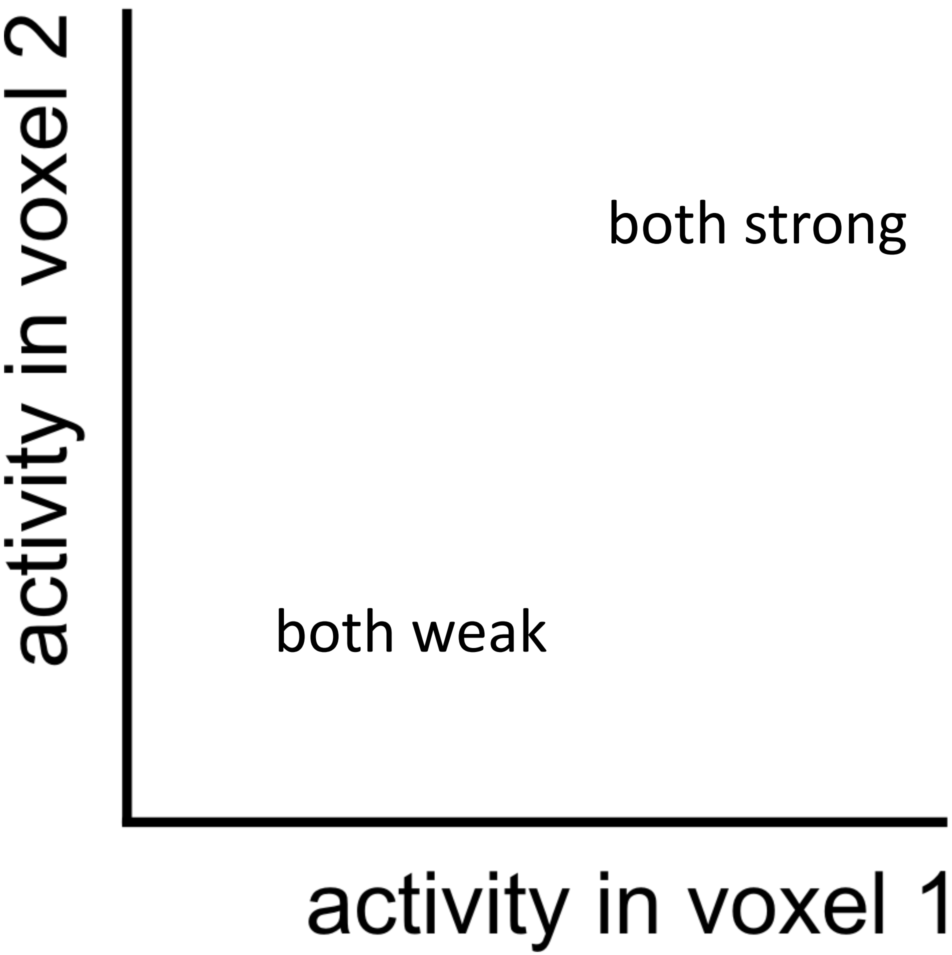
Overview

- Why classification analysis?
- **Linear classification: the basic idea**
- Linear classification: different classifiers
- Do it yourself: six steps
  - step 1: split data and preprocess
  - step 2: estimate single-subject activity patterns
  - step 3: select voxels
  - step 4: train the classifier
  - step 5: test the classifier
  - step 6: statistical inference
- Toolboxes
- Literature

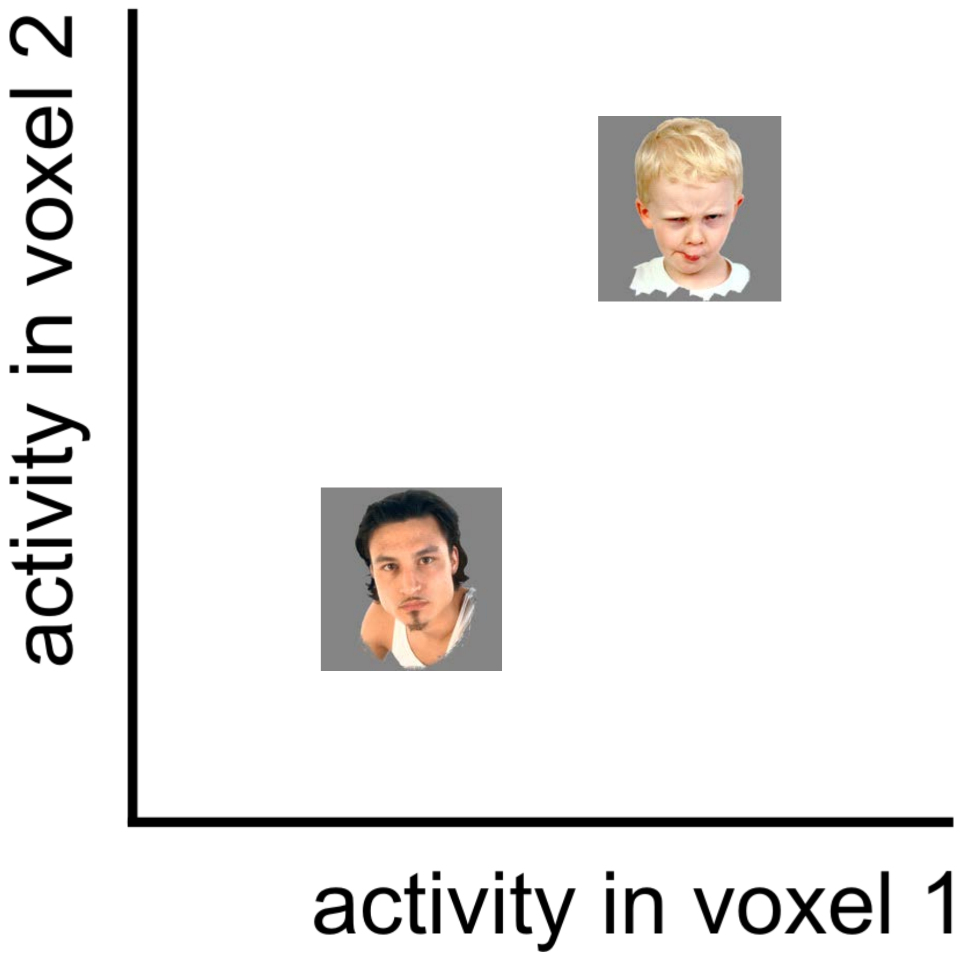
Linear classification: the basic idea



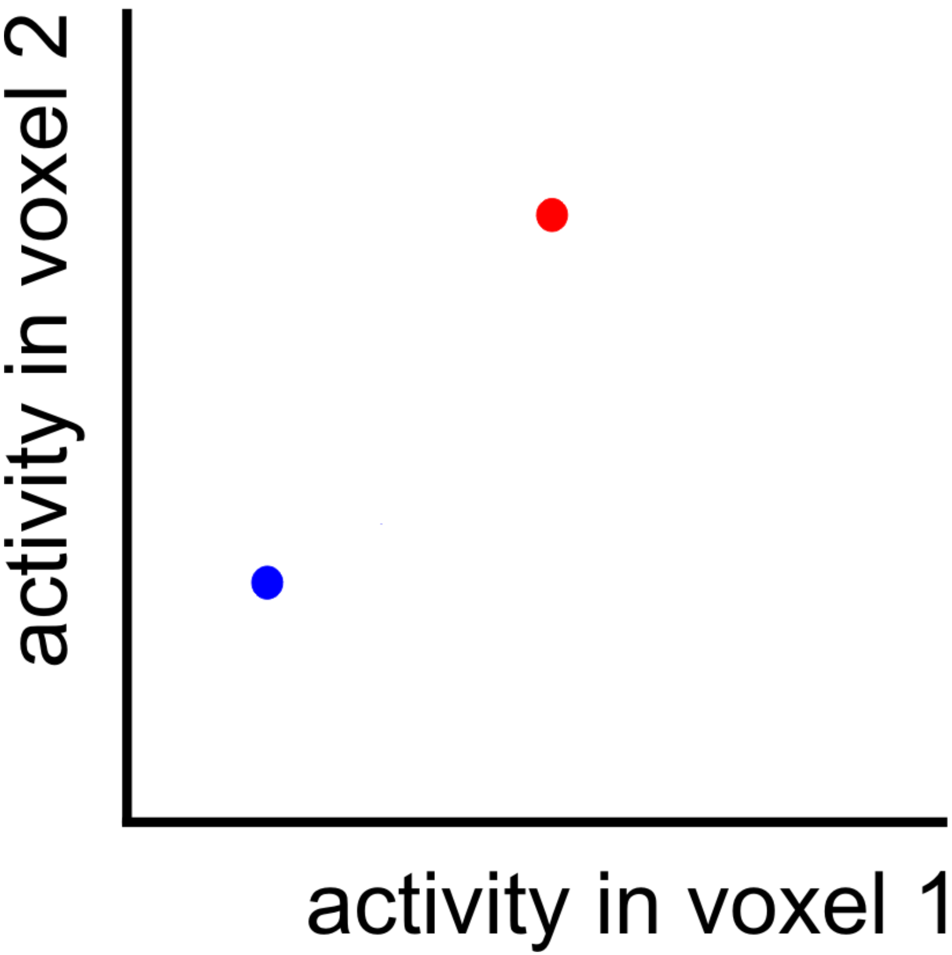
Linear classification: the basic idea



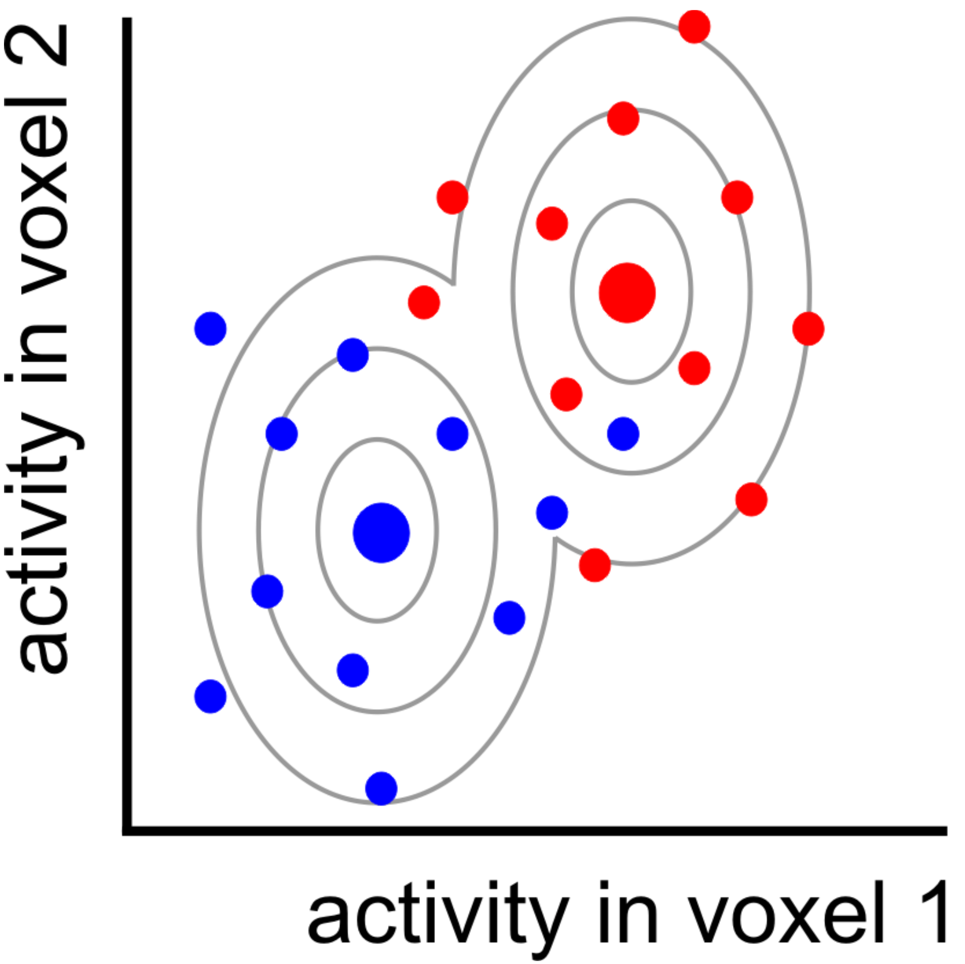
Linear classification: the basic idea



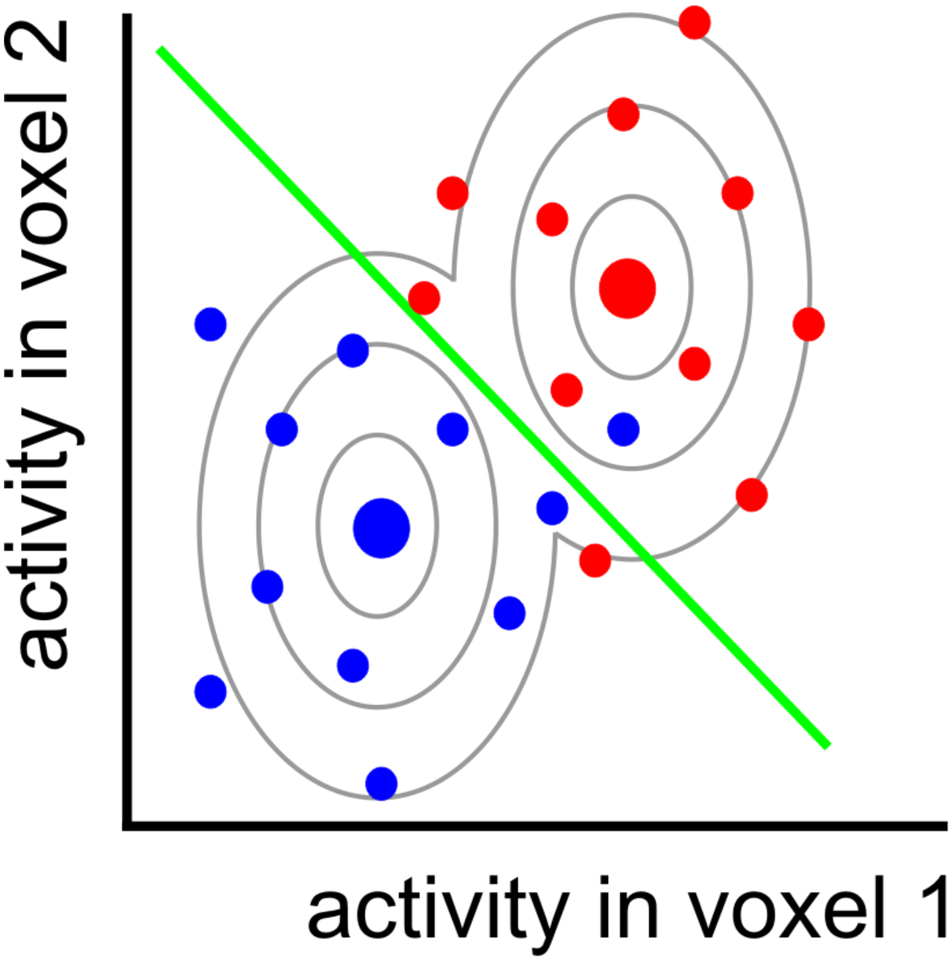
Linear classification: the basic idea

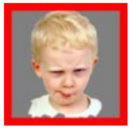


Linear classification: the basic idea



Linear classification: the basic idea

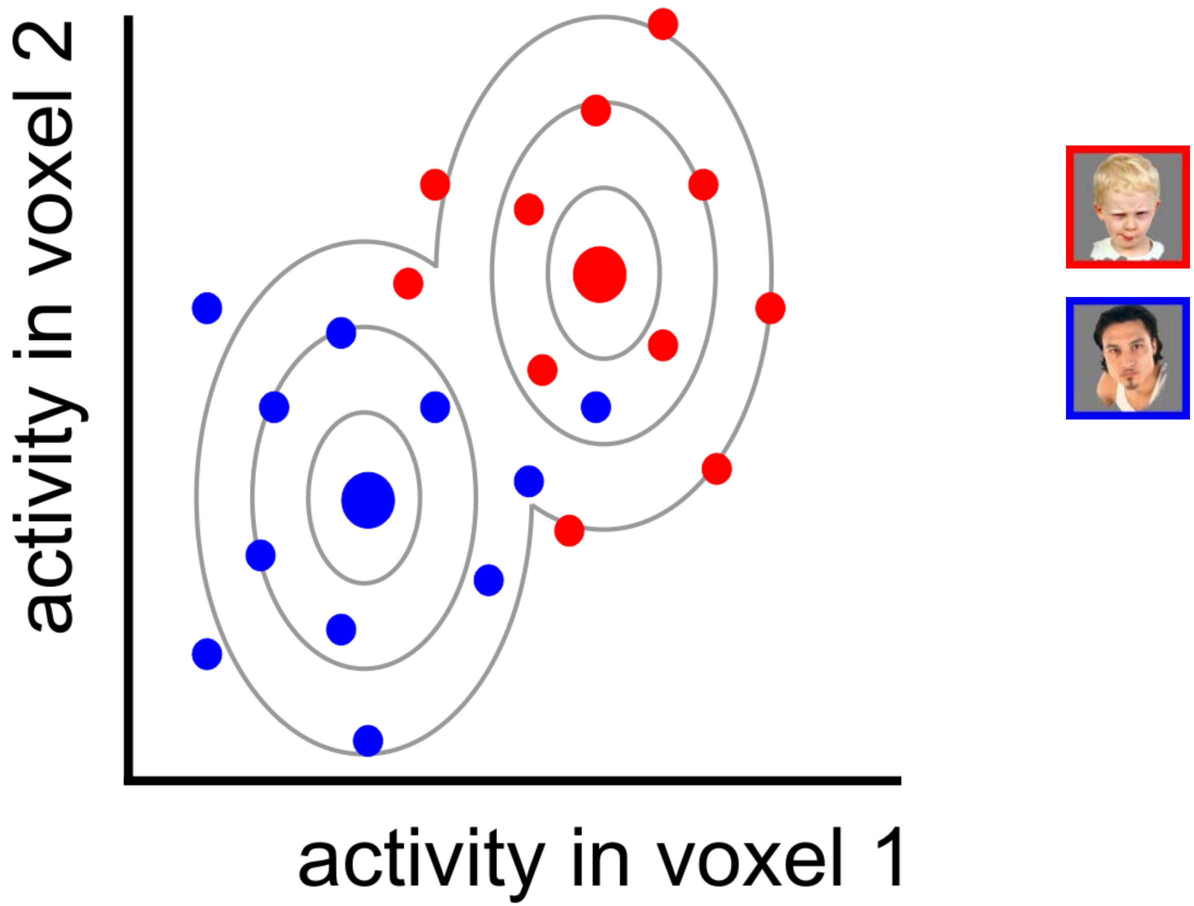




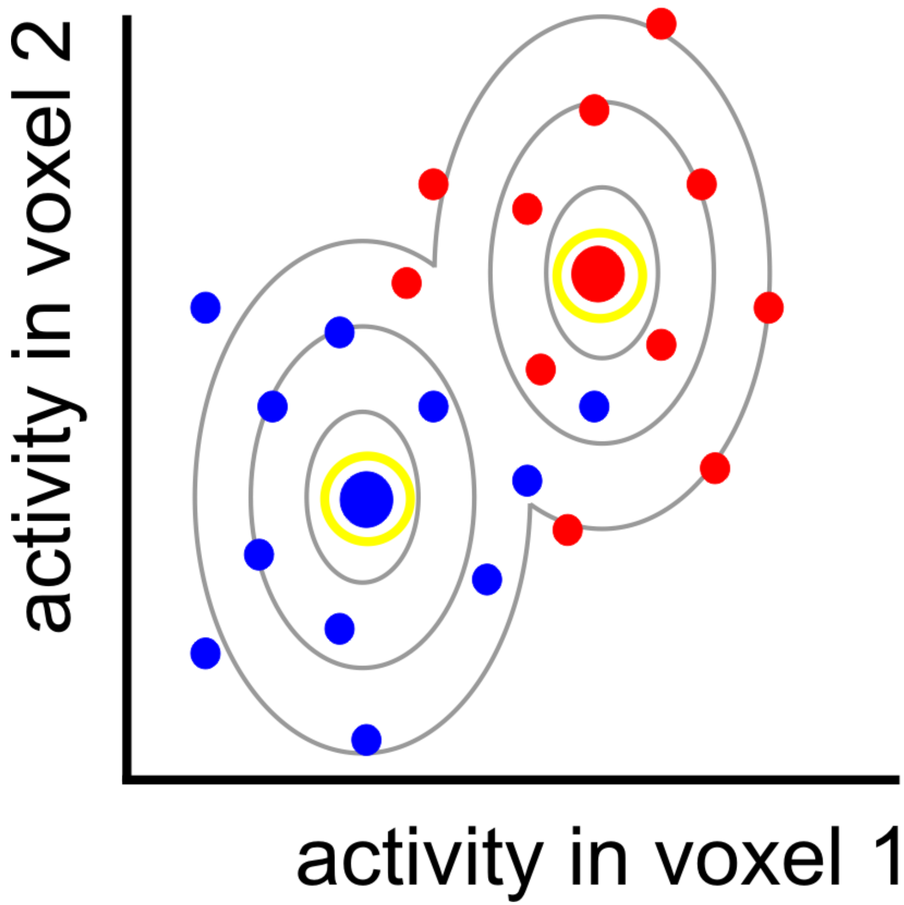
Overview

- Why classification analysis?
- Linear classification: the basic idea
- **Linear classification: different classifiers**
- Do it yourself: six steps
  - step 1: split data and preprocess
  - step 2: estimate single-subject activity patterns
  - step 3: select voxels
  - step 4: train the classifier
  - step 5: test the classifier
  - step 6: statistical inference
- Toolboxes
- Literature

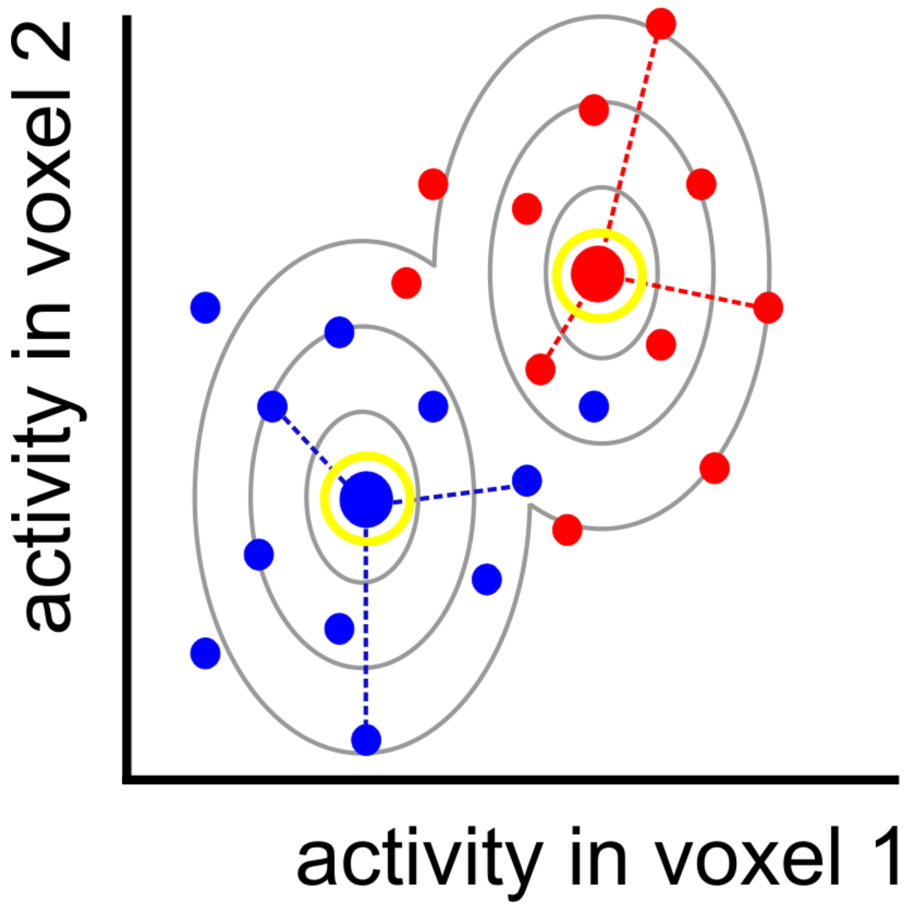
Linear classification: different classifiers



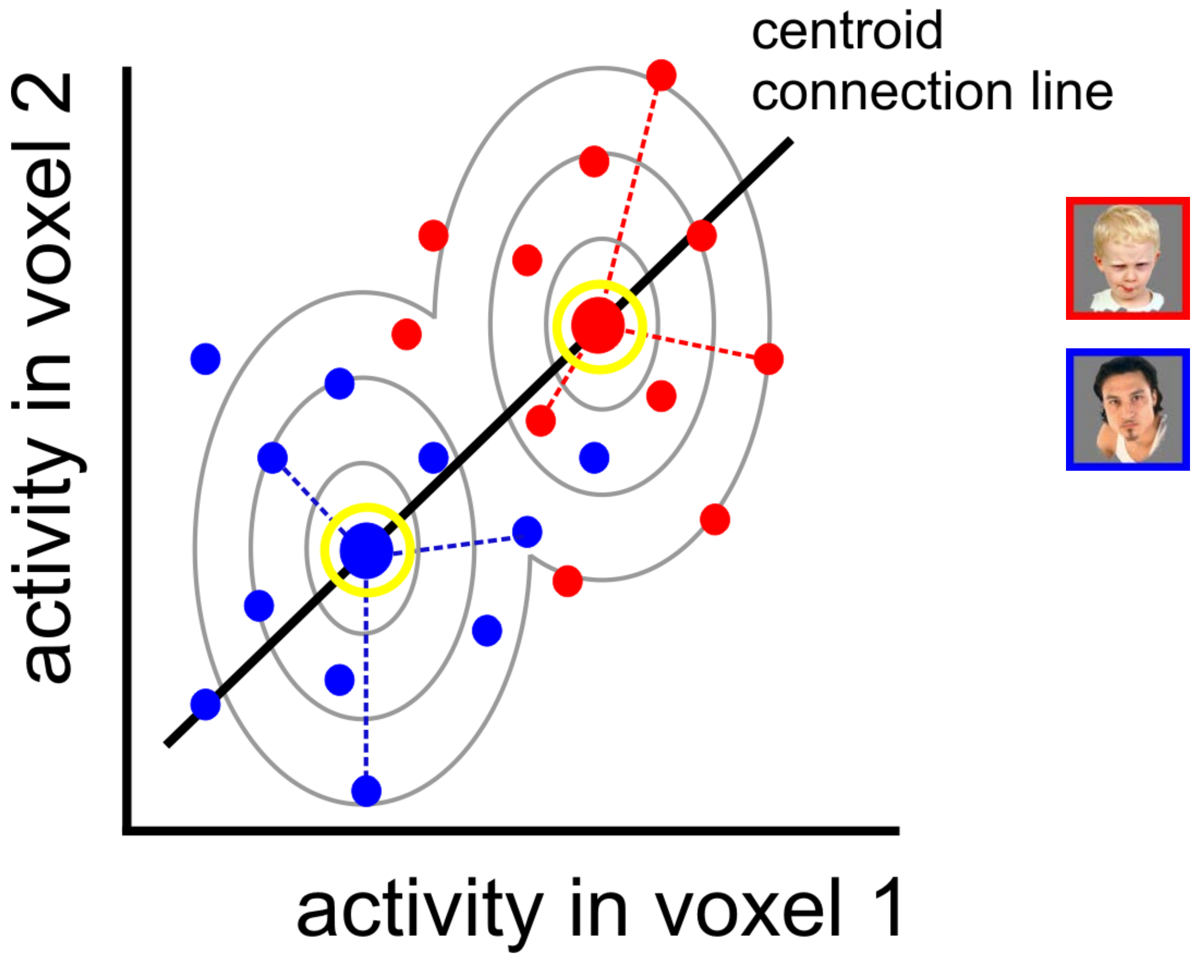
Linear classification: minimum-distance



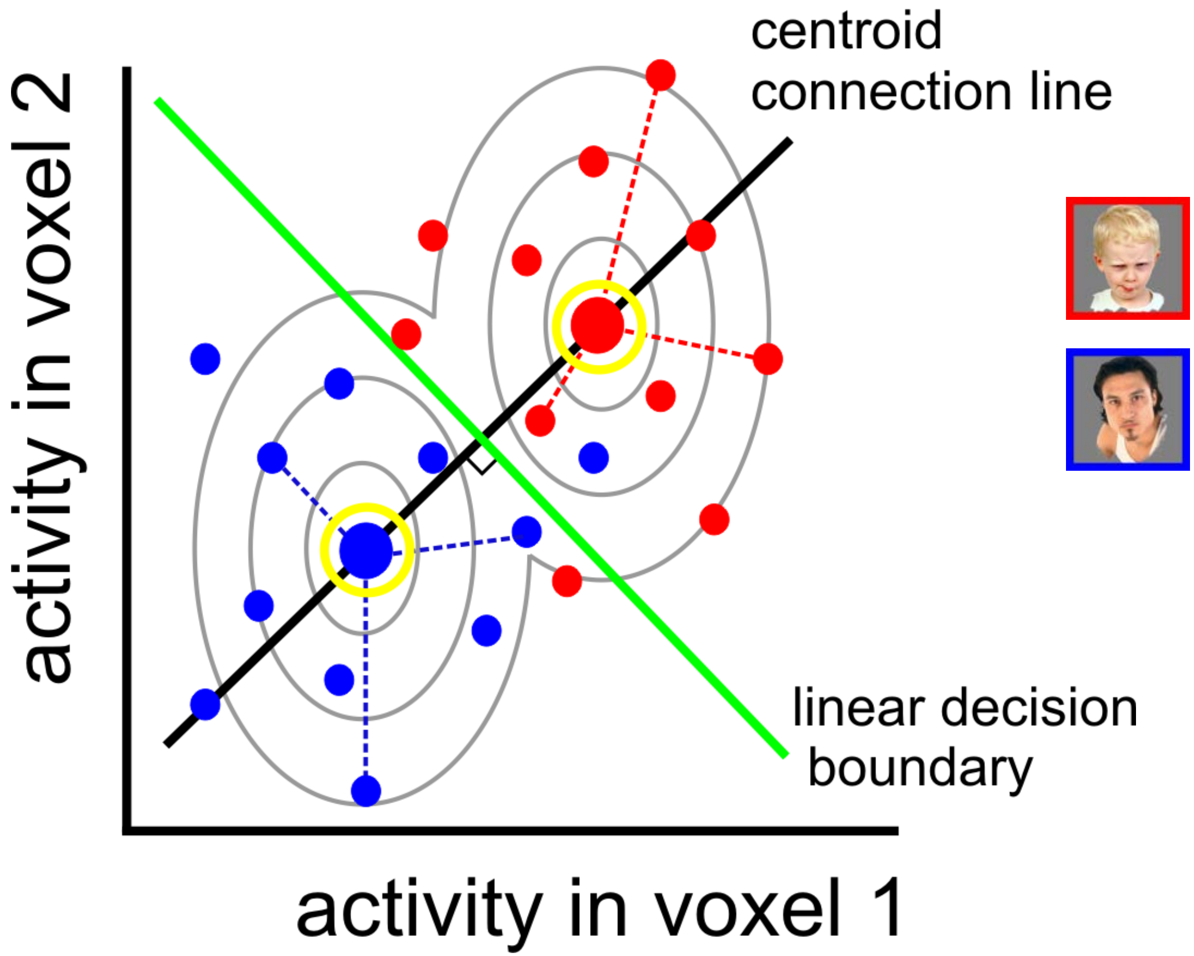
Linear classification: minimum-distance



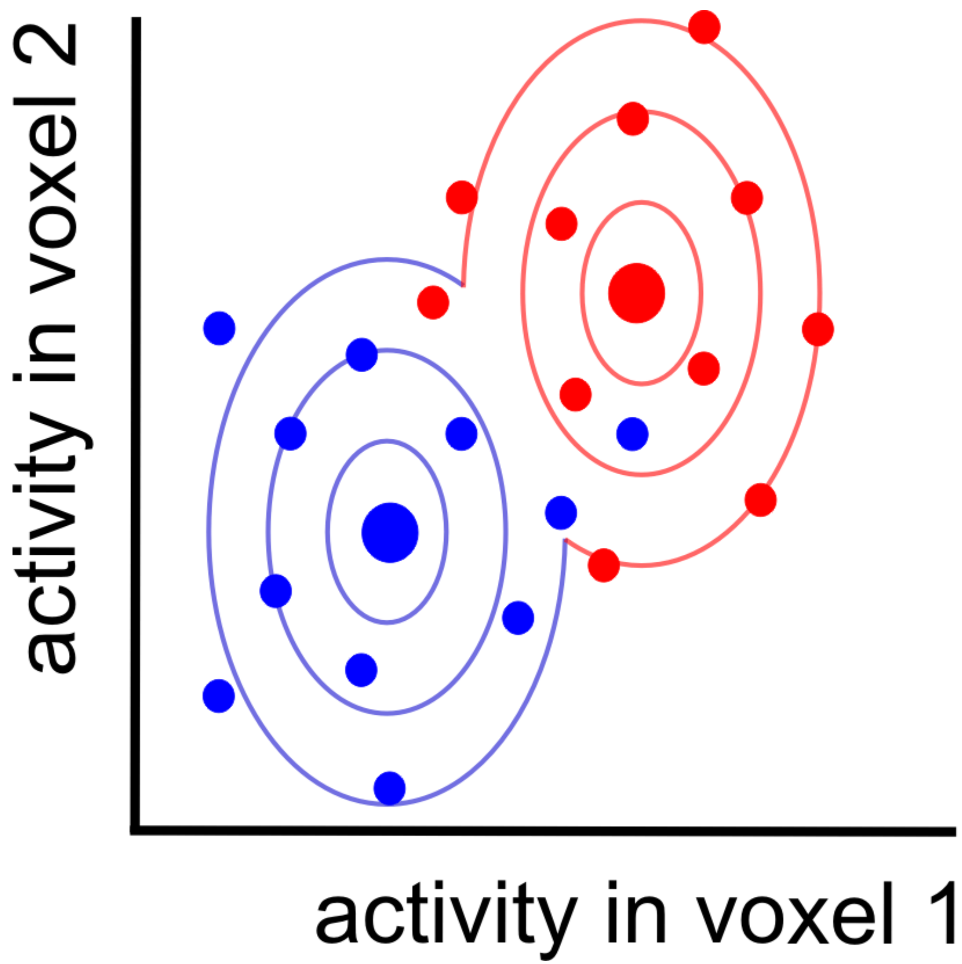
Linear classification: minimum-distance



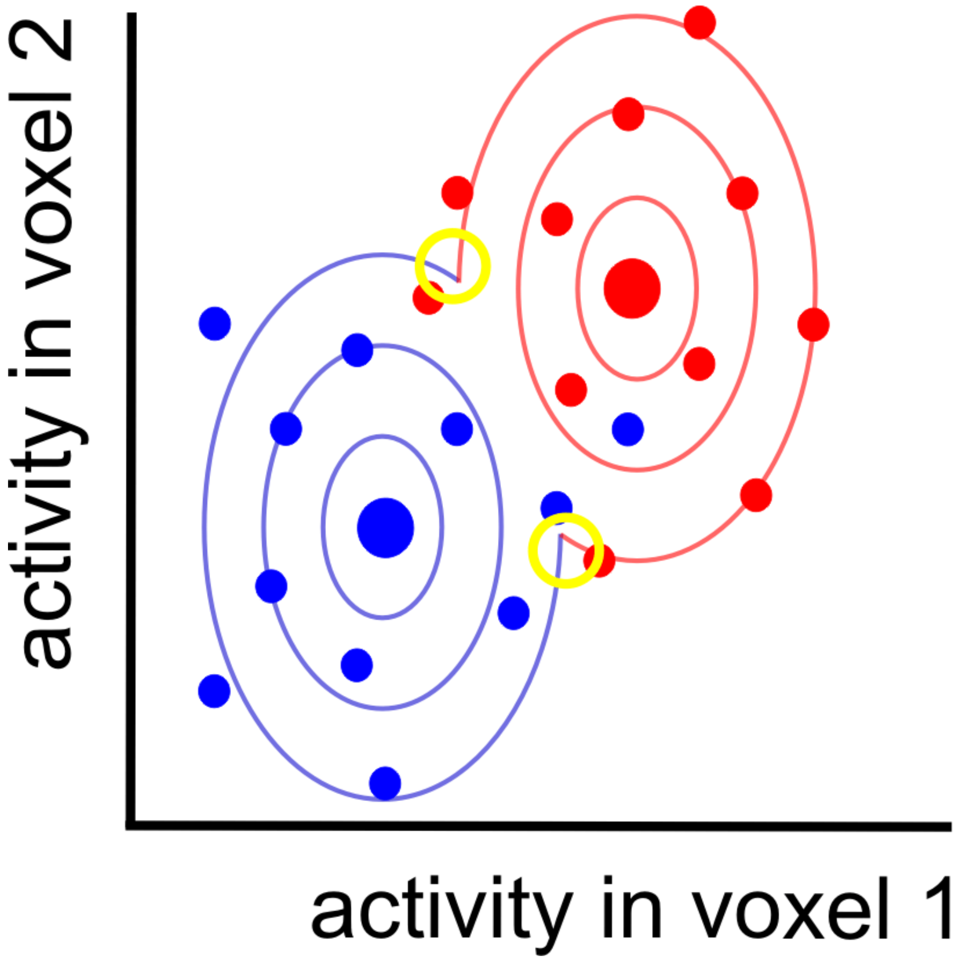
Linear classification: minimum-distance



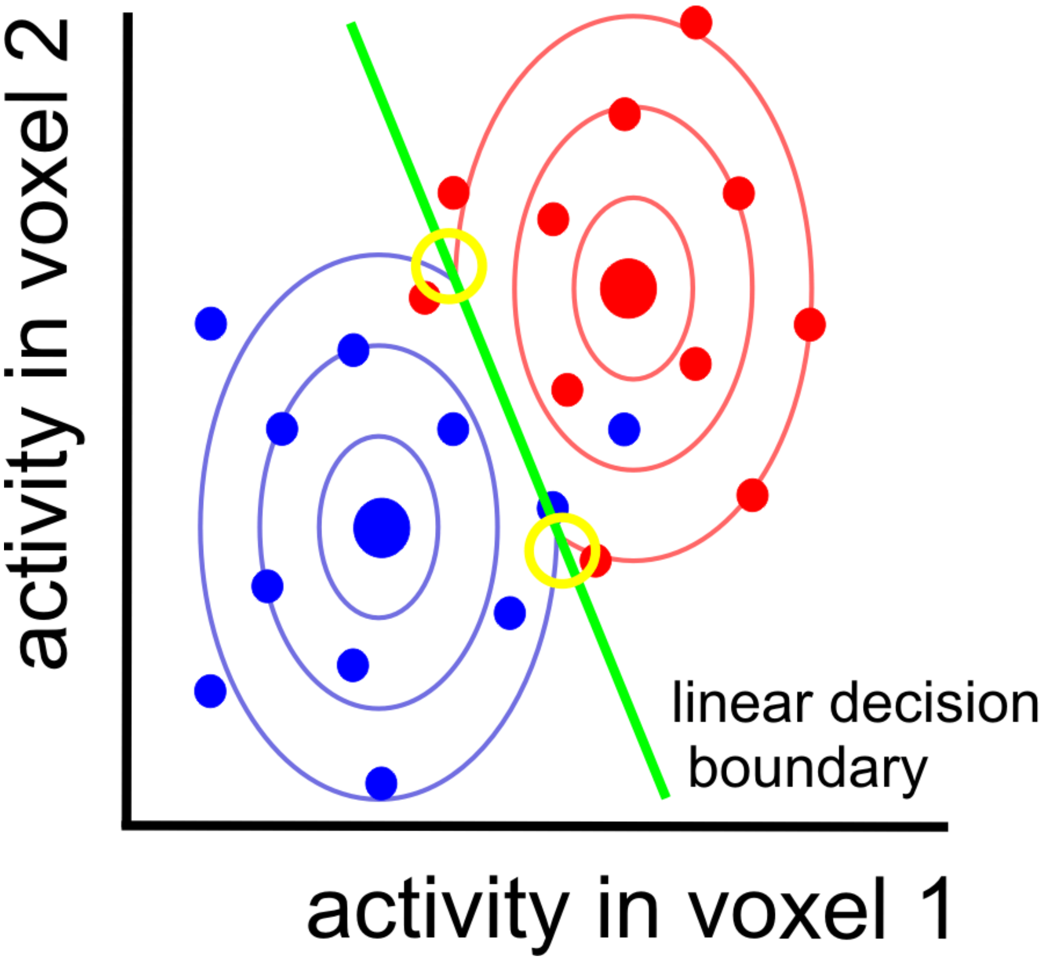
Linear classification: FLDA



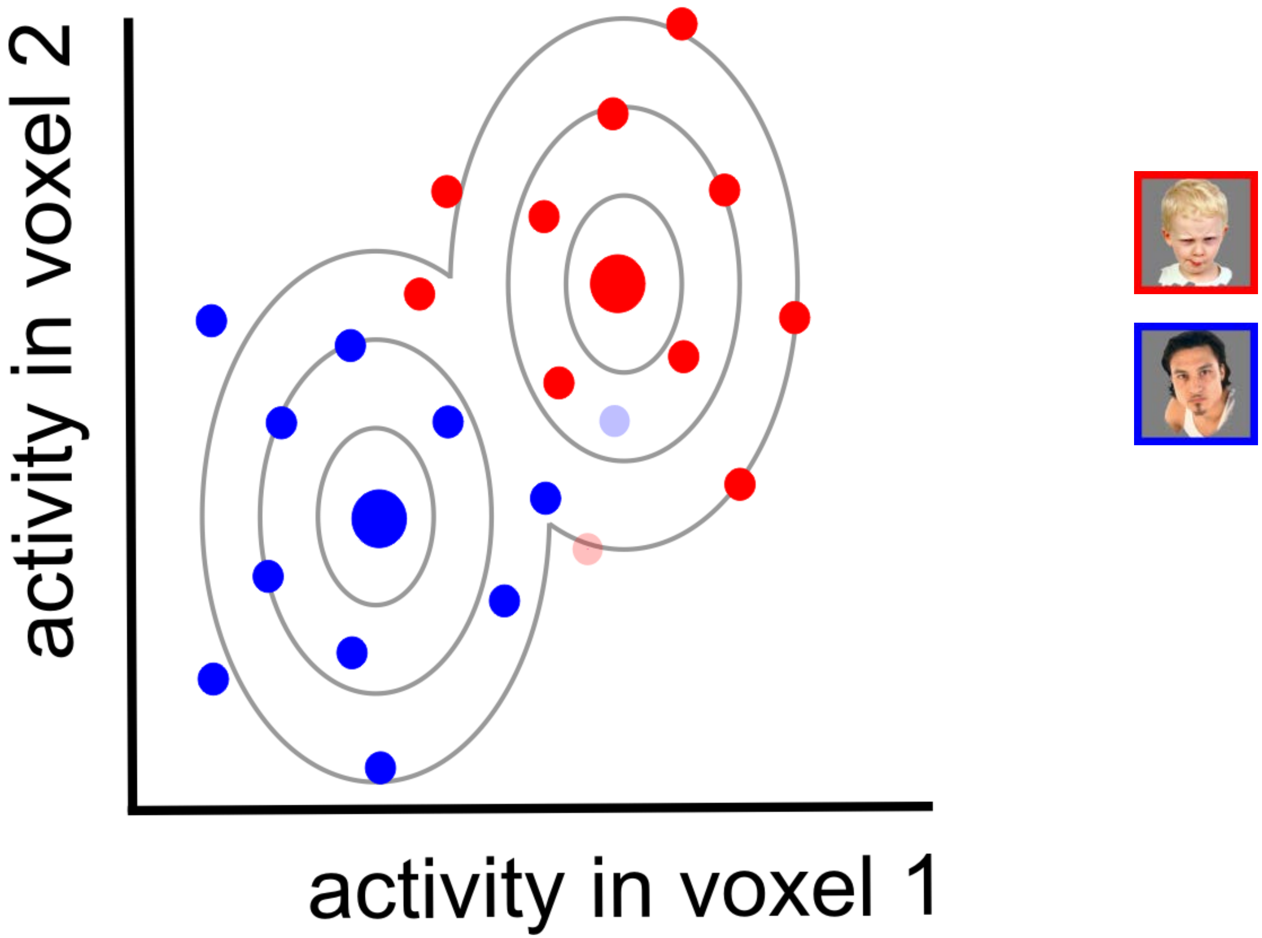
Linear classification: FLDA



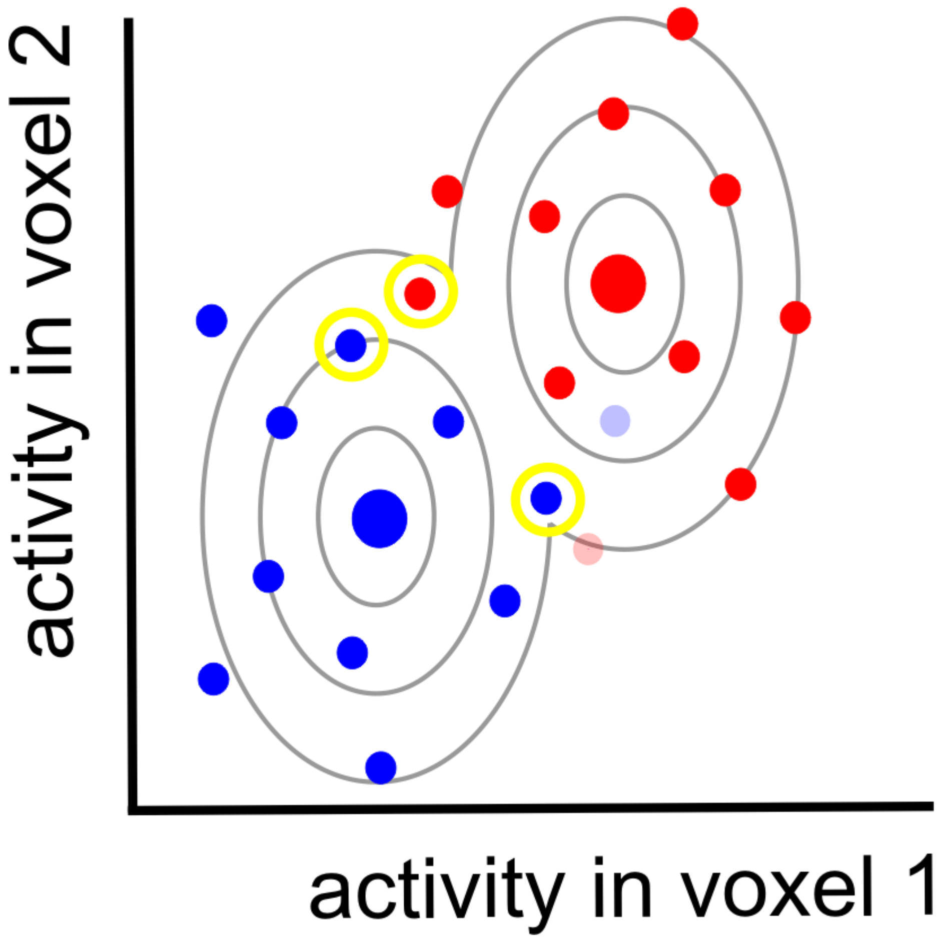
Linear classification: FLDA

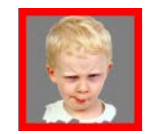


Linear classification: linear SVM

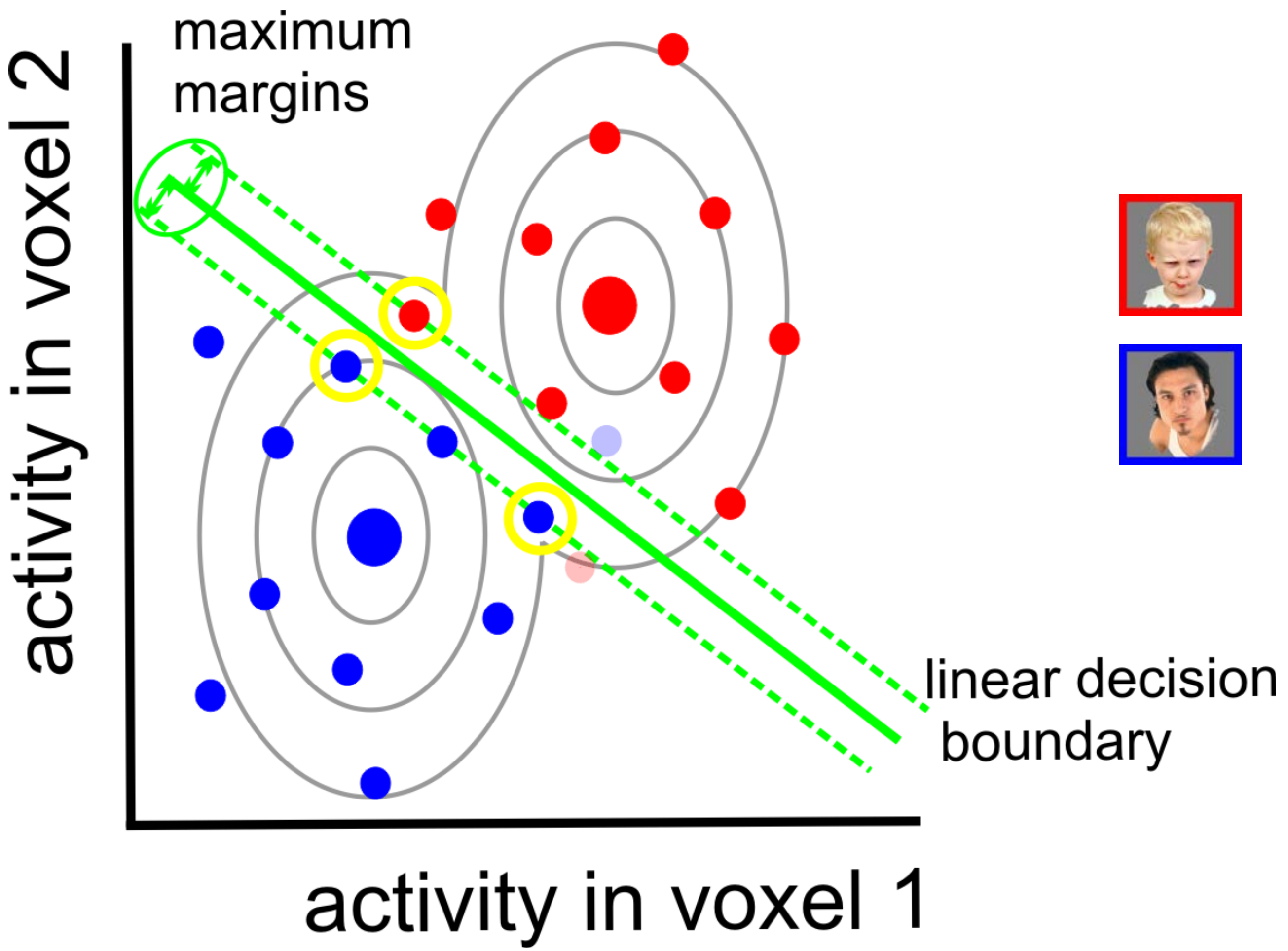


Linear classification: linear SVM

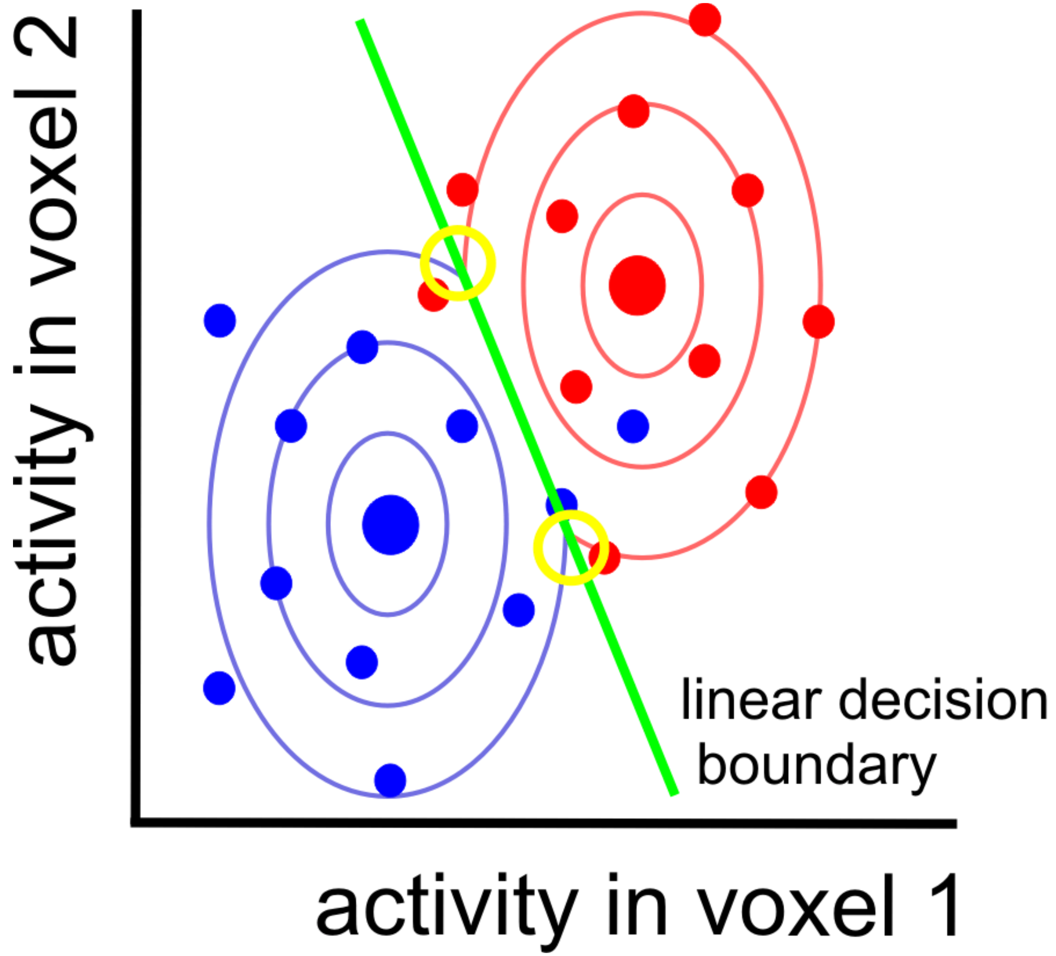




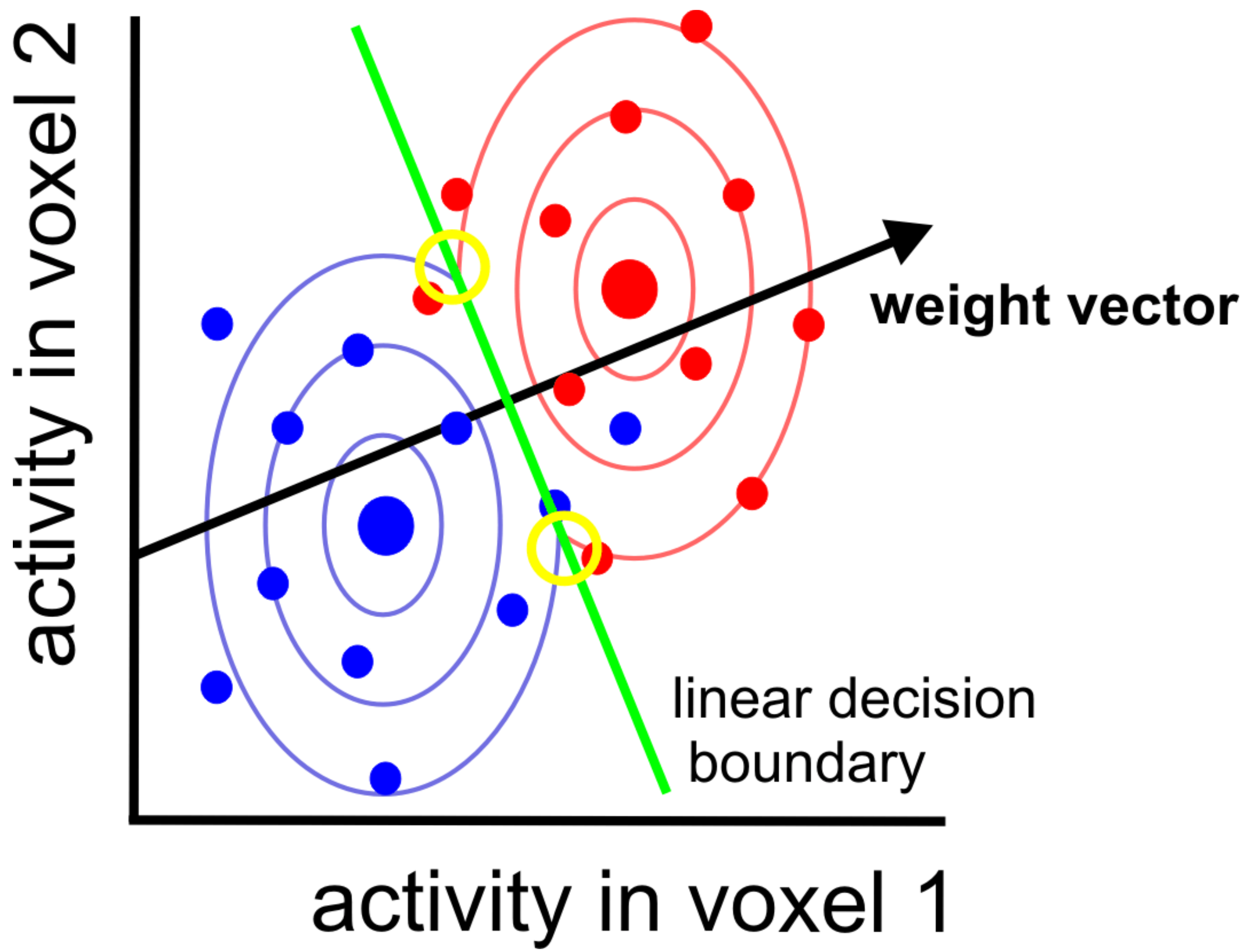
Linear classification: linear SVM



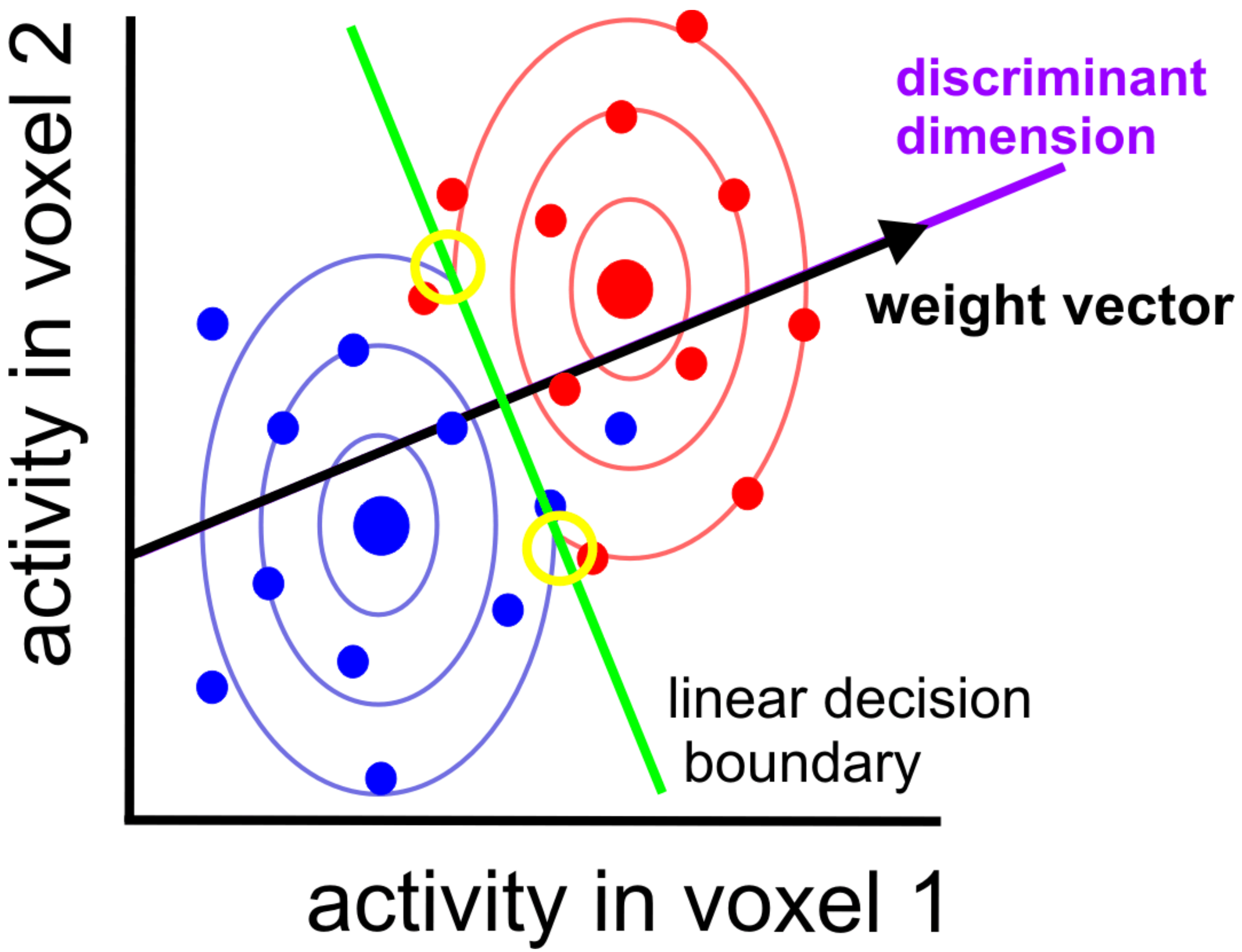
Boundary placement: some more detail



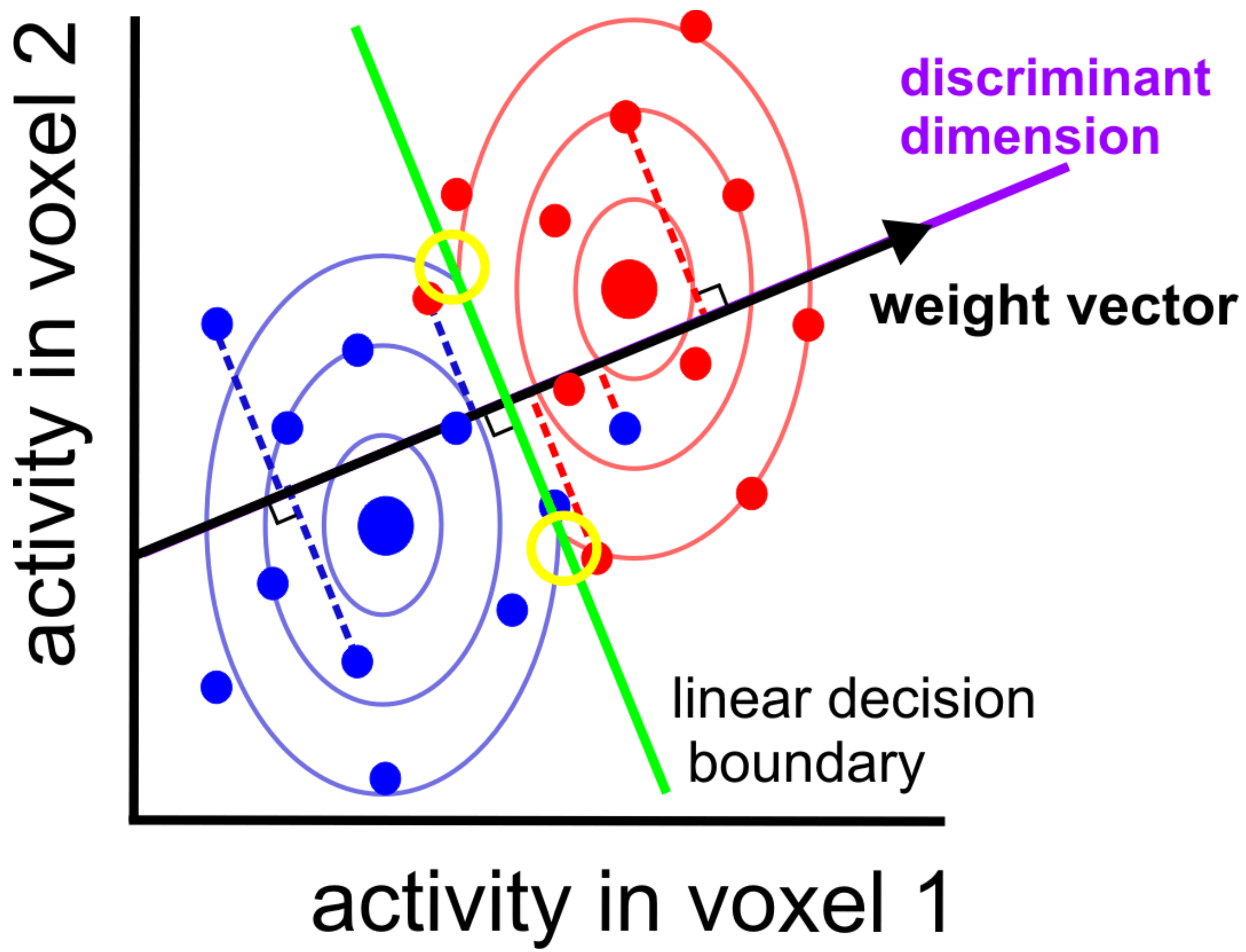
Boundary placement: some more detail



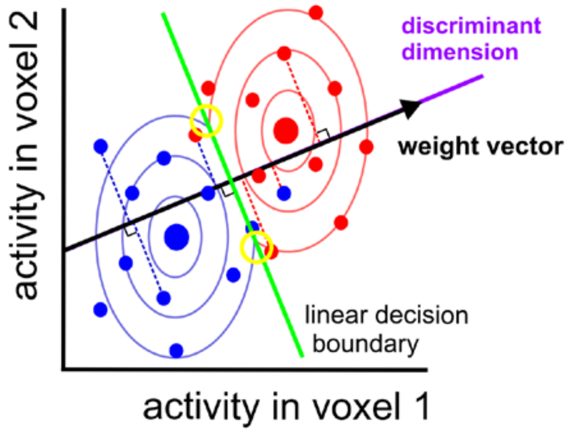
Boundary placement: some more detail



Boundary placement: some more detail



Boundary placement: some more detail



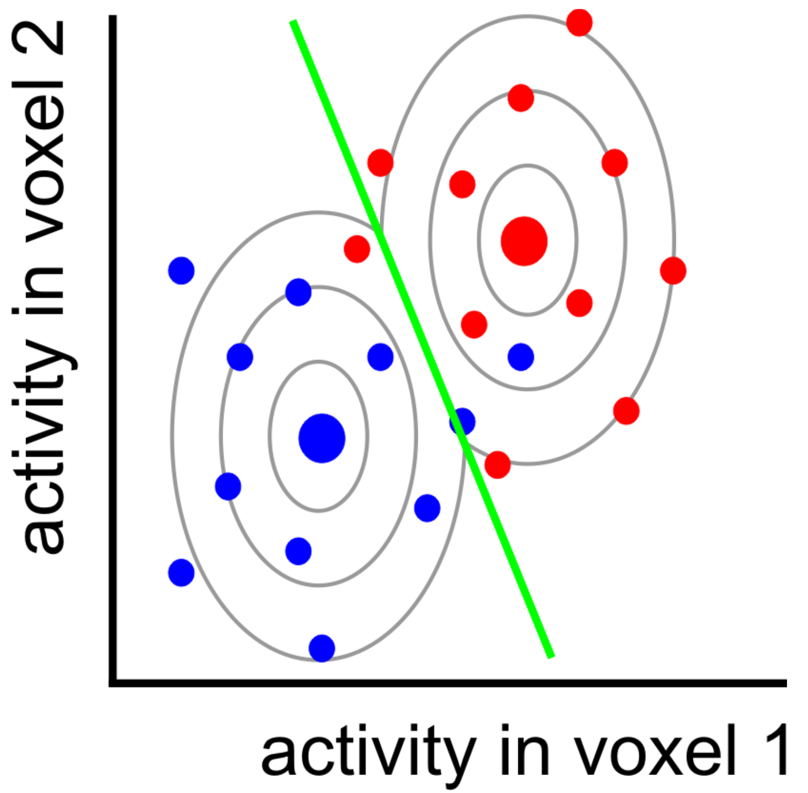
Fisher linear discriminant

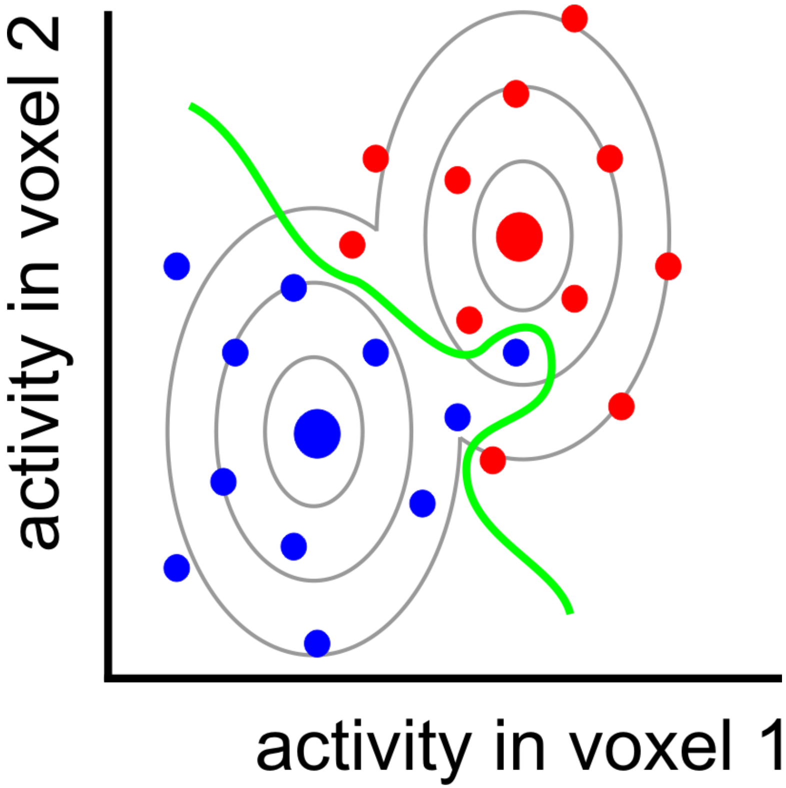
assumes identical and isotropic distributions

assumes identical and multivariate normal distributions

no assumptions about distributions

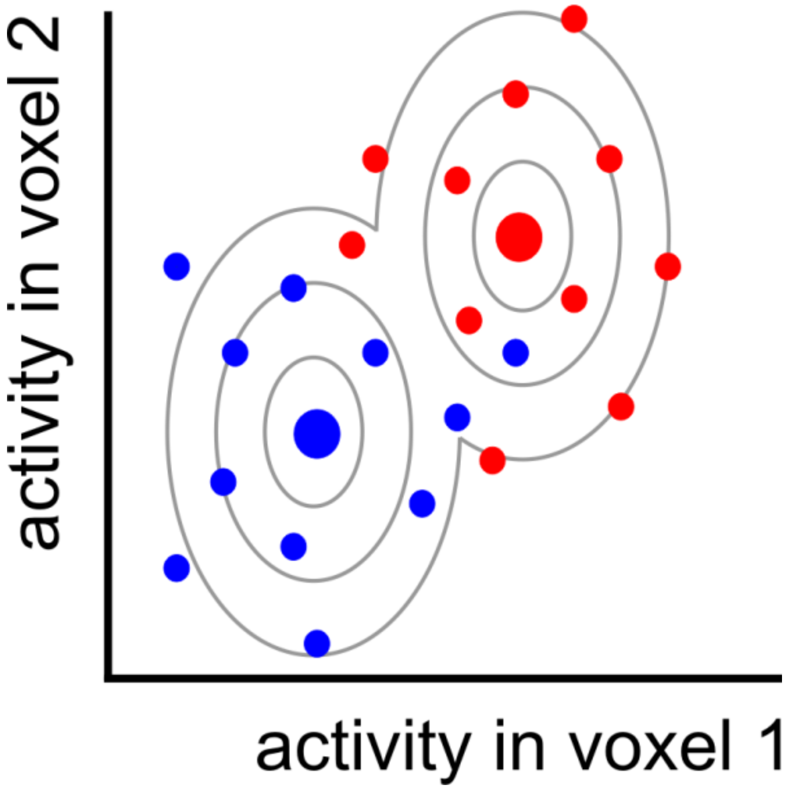
Can we do better?



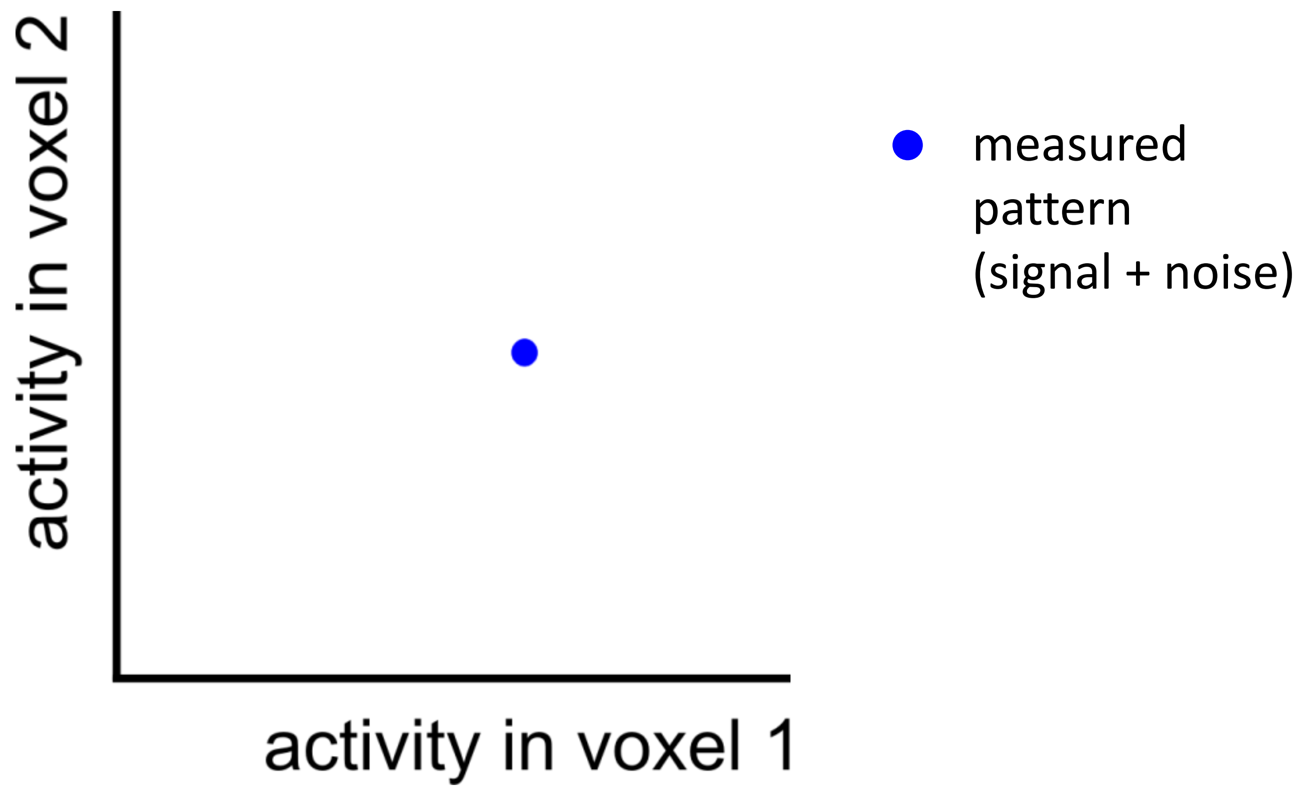


nonlinear
classifier

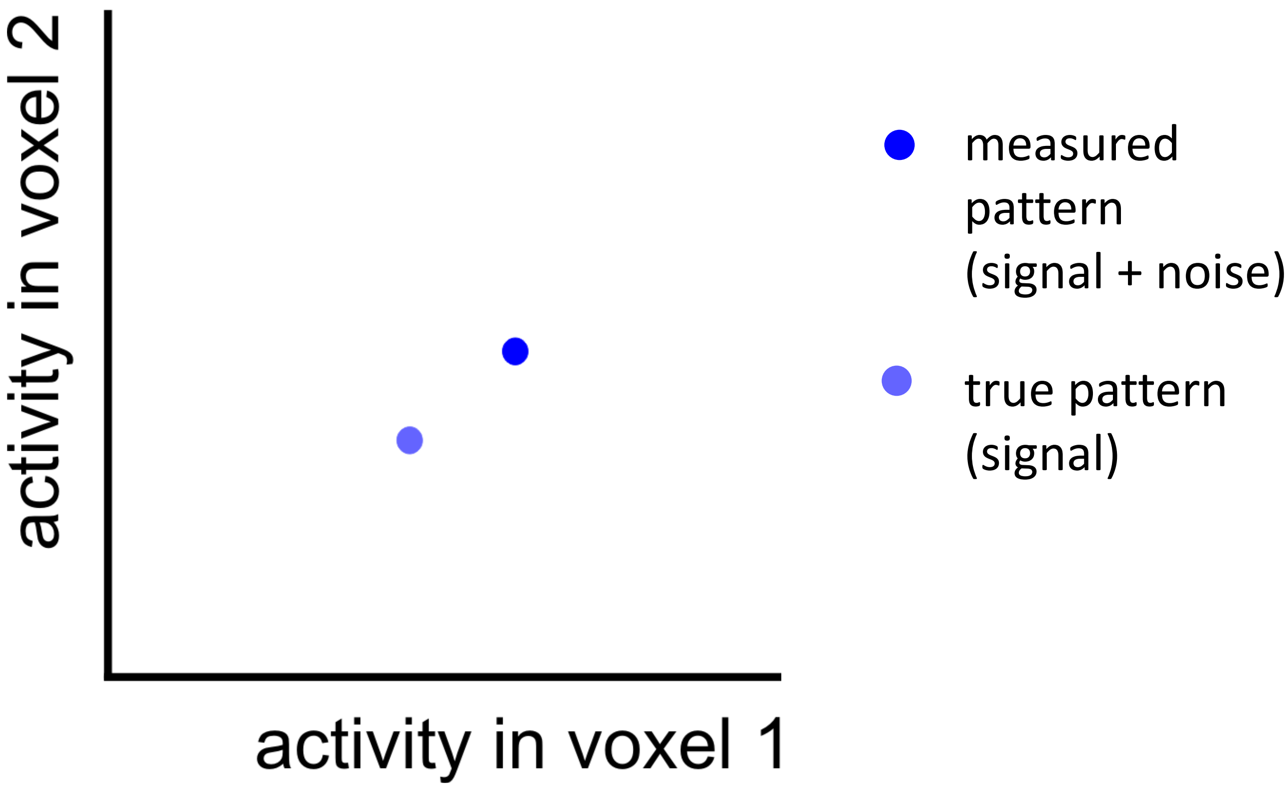
Pattern = signal + noise



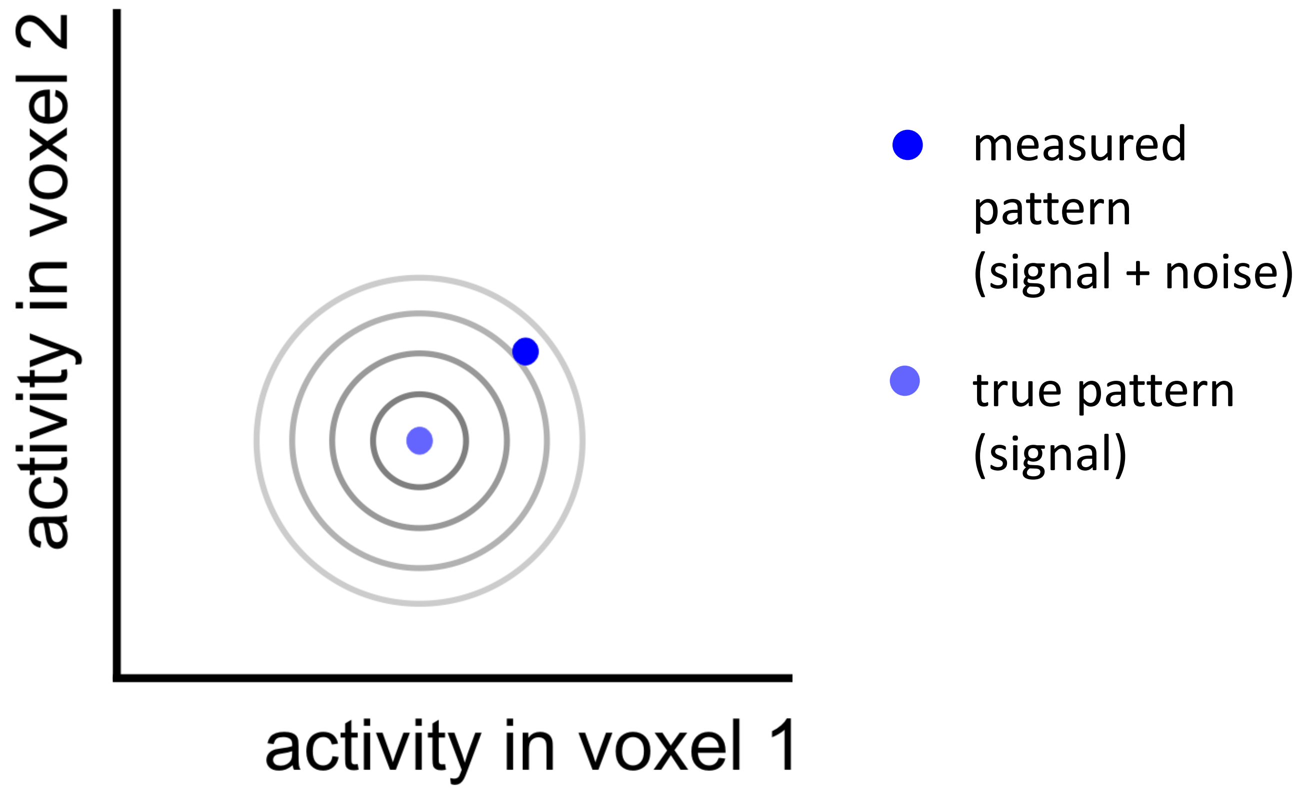
Pattern = signal + noise



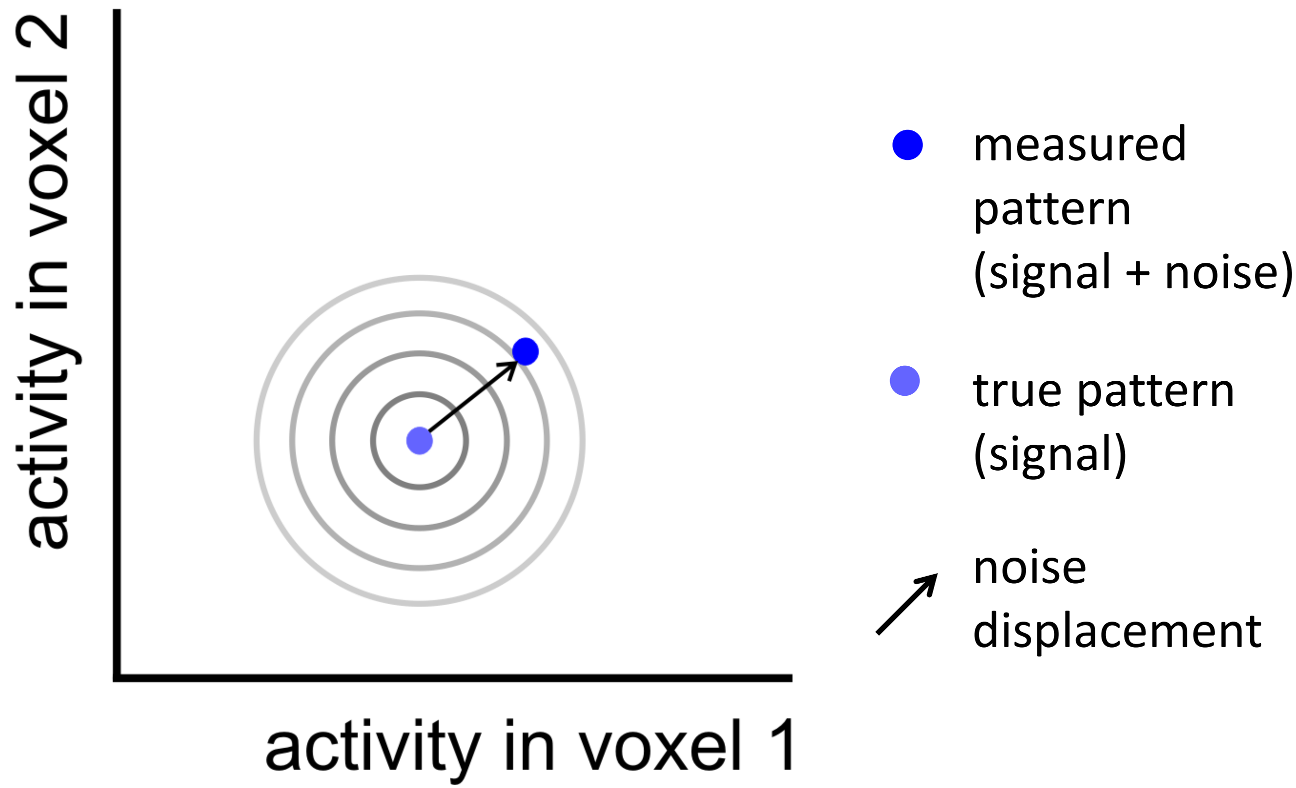
Pattern = signal + noise



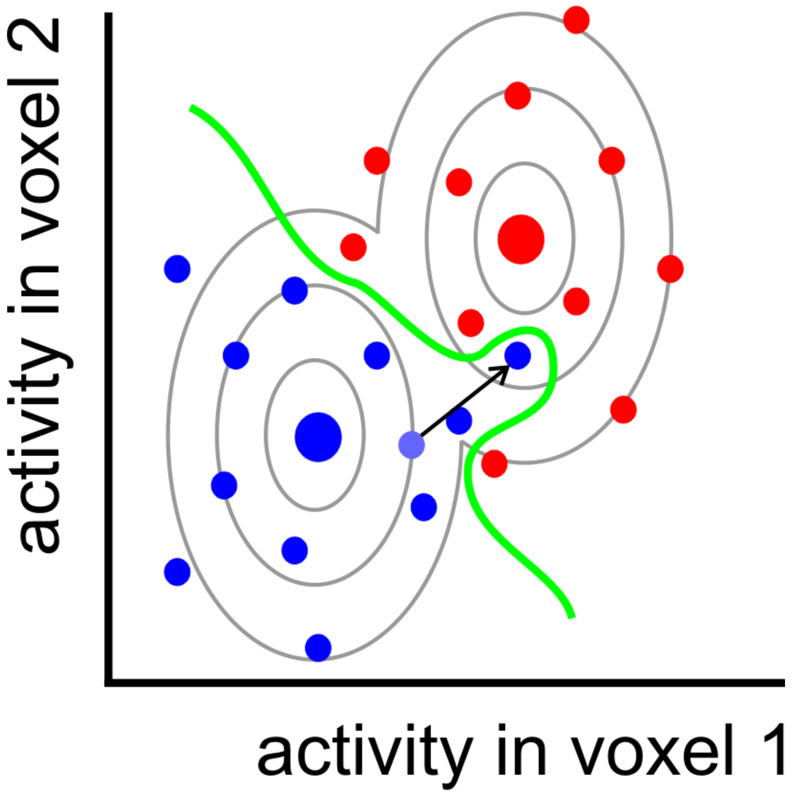
Pattern = signal + noise



Pattern = signal + noise

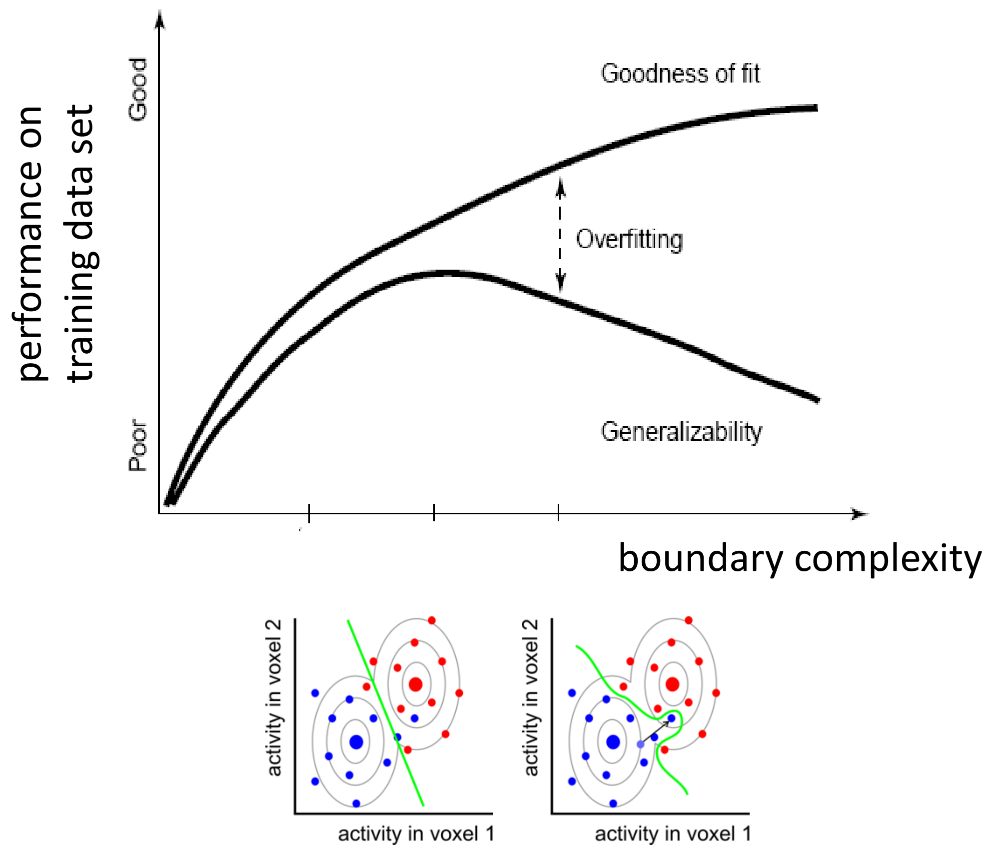


Overfitting



Overfitting

After determining the decision boundary, we need to test how well the boundary generalises to new data (cross validation).



Overfitting

After determining the decision boundary, we need to test how well the boundary generalises to new data (cross validation).

Linear classifiers usually perform better on fMRI data than nonlinear classifiers.

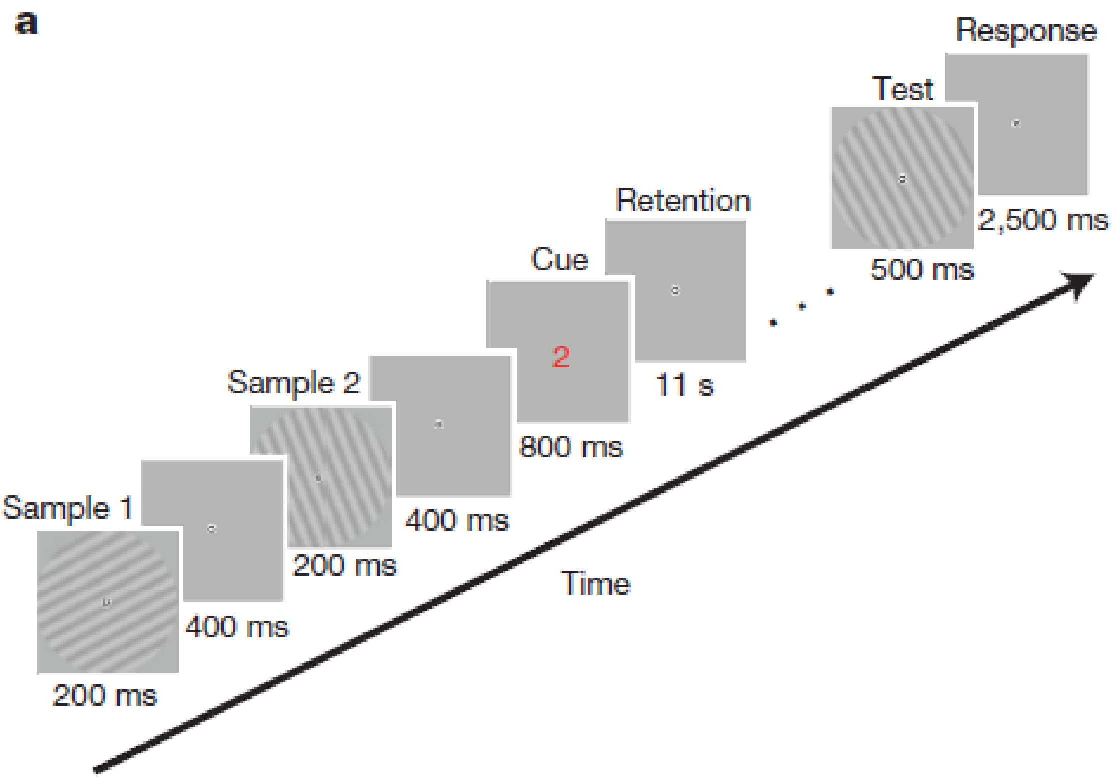
Overfitting can be further reduced by:

- regularisation
- dimensionality reduction of the activity patterns (e.g. voxel selection)

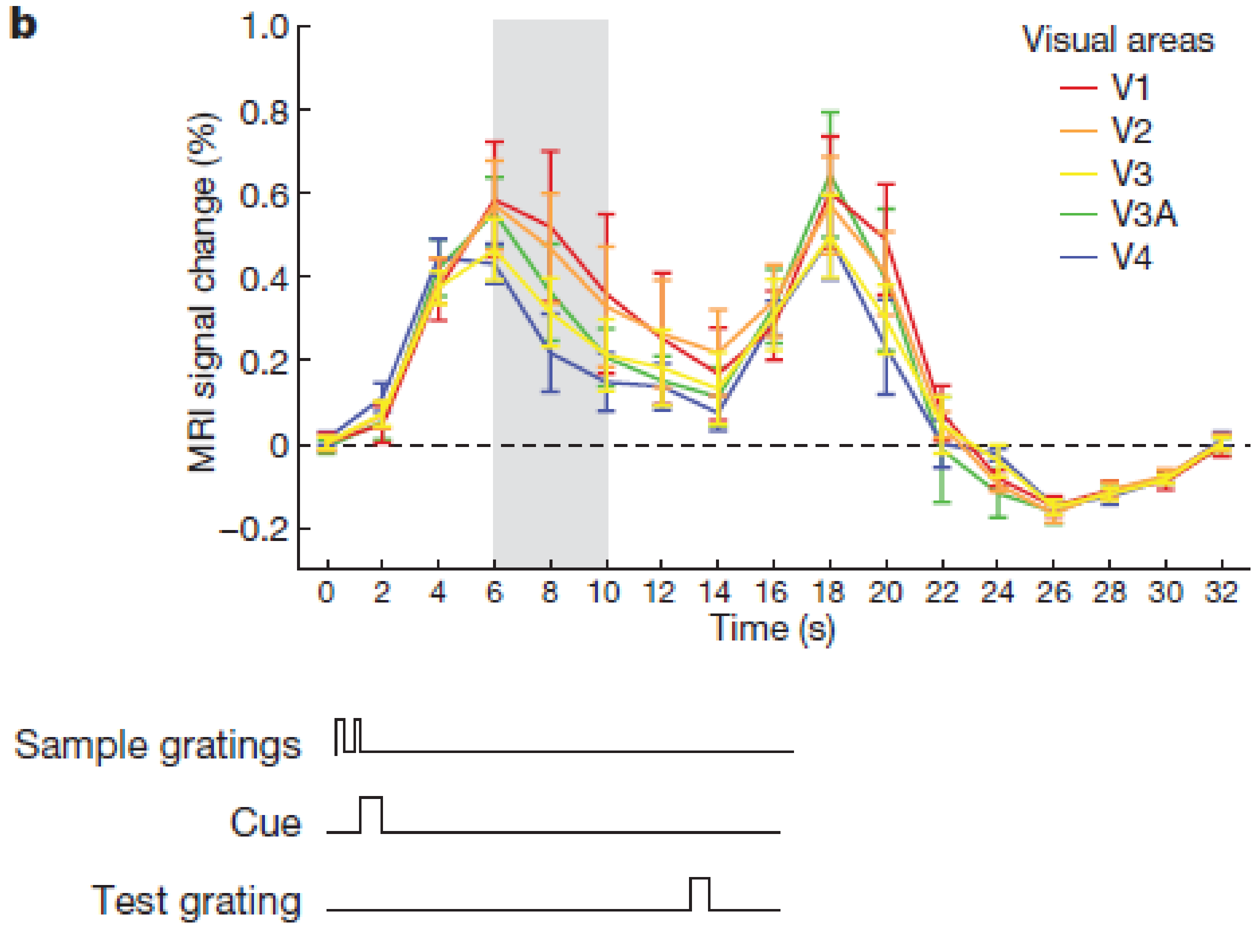
Applications: visual WM

Decoding reveals the contents of visual working memory in early visual areas

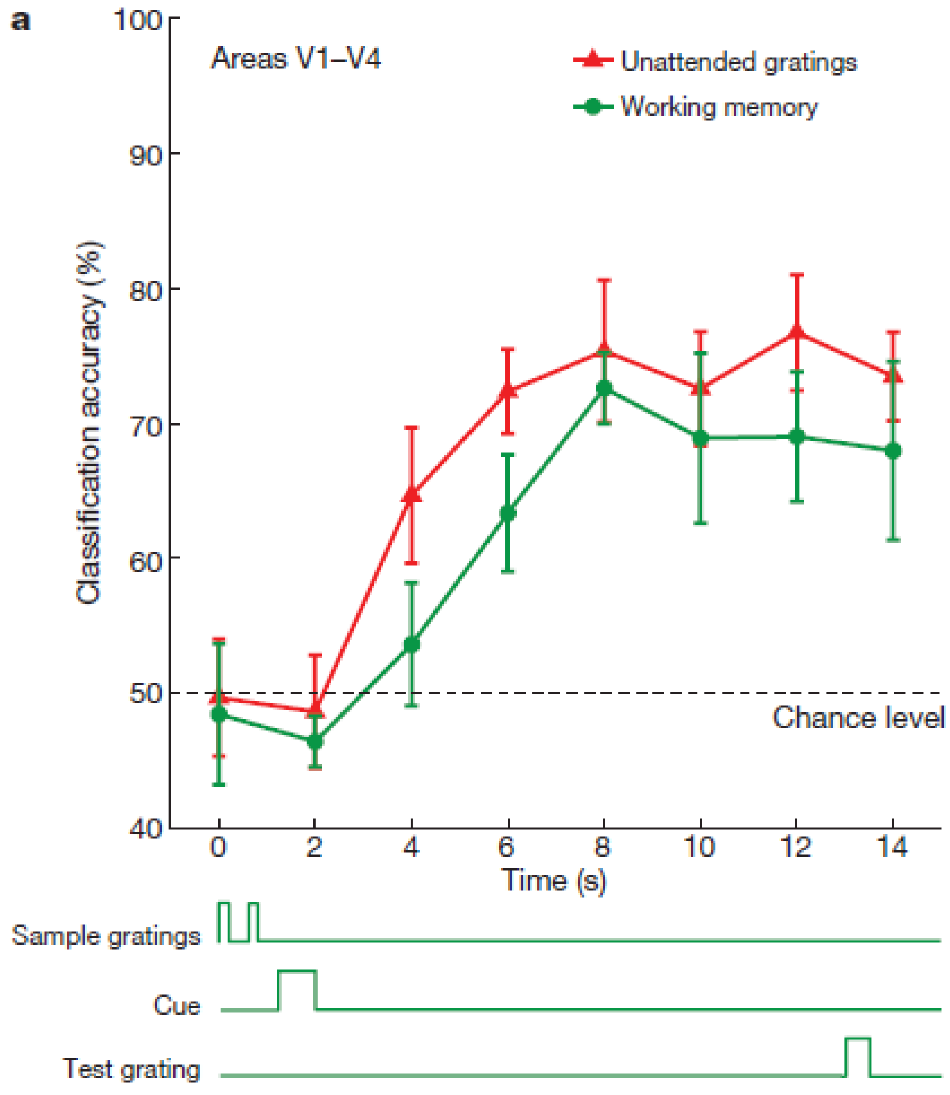
Stephenie A. Harrison[1] & Frank Tong[1]



Applications: visual WM



Applications: visual WM



Overview

- Why classification analysis?
- Linear classification: the basic idea
- Linear classification: different classifiers
- **Do it yourself: six steps**
  - step 1: split data and preprocess
  - step 2: estimate single-subject activity patterns
  - step 3: select voxels
  - step 4: train the classifier
  - step 5: test the classifier
  - step 6: statistical inference
- Toolboxes
- Literature

Do it yourself: six steps

Step 1: split data and preprocess

Step 2: estimate single-subject activity patterns

Step 3: select voxels

Step 4: train the classifier

Step 5: test the classifier

Step 6: statistical inference

Do it yourself: six steps

Step 1: split data and preprocess

Step 2: estimate single-subject activity patterns

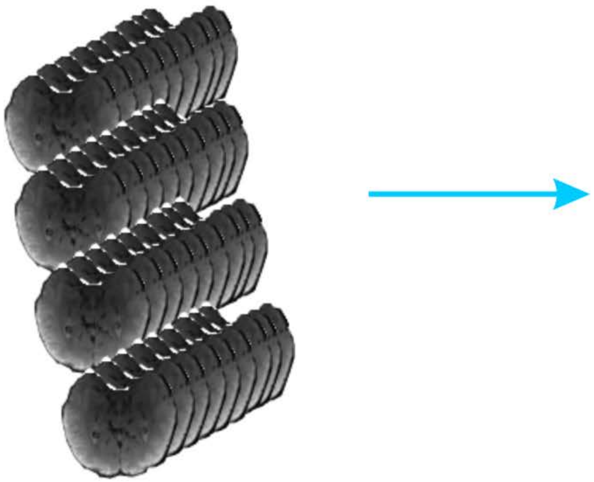
Step 3: select voxels

Step 4: train the classifier

Step 5: test the classifier

Step 6: statistical inference

Step 1a: split data



full data set

Make sure that training and test data are independent.

Step 1b: preprocess

As usual:

- slice-scan-time correction
- motion-correction

Optional:

- normalisation to template (if random-effects searchlight analysis across subjects)
- spatial smoothing (to increase signal, sensitive to larger-scale spatial patterns)

Do it yourself: six steps

Step 1: split data and preprocess

Step 2: estimate single-subject activity patterns

Step 3: select voxels

Step 4: train the classifier

Step 5: test the classifier

Step 6: statistical inference

Step 2: estimate single-subject activity patterns

t patterns preferred over beta patterns (Misaki et al. 2010)

Do it yourself: six steps

Step 1: split data and preprocess

Step 2: estimate single-subject activity patterns

Step 3: select voxels

Step 4: train the classifier

Step 5: test the classifier

Step 6: statistical inference

Step 3: select voxels

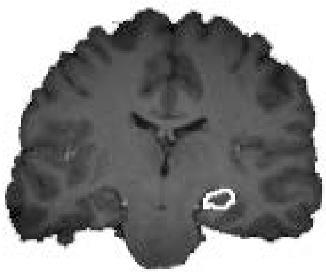
Make sure that voxel selection is based on data independent from test data set.

Most common ways of voxel selection:

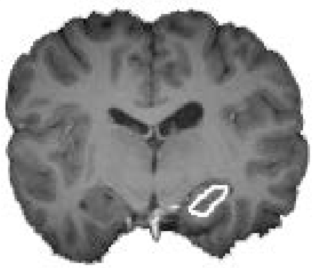
- structural selection (anatomy)
- functional selection (activity)
  - univariate (activation differences)
  - multivariate (pattern differences)
- geometrical selection
  - multivoxel searchlight

Step 3: select voxels

anatomy



subject 1



subject 2

•

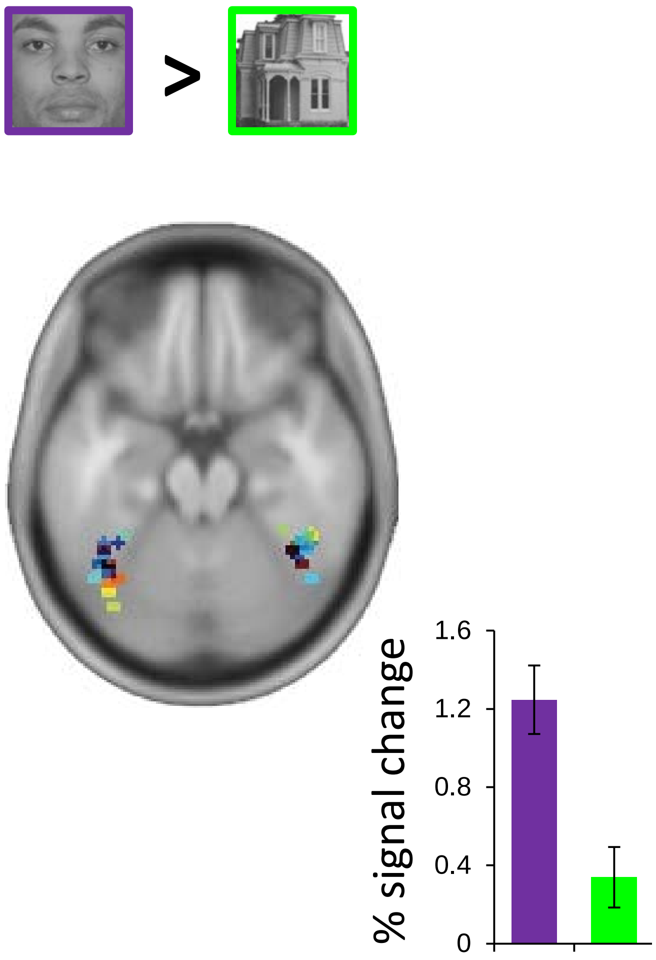
•

• subject n

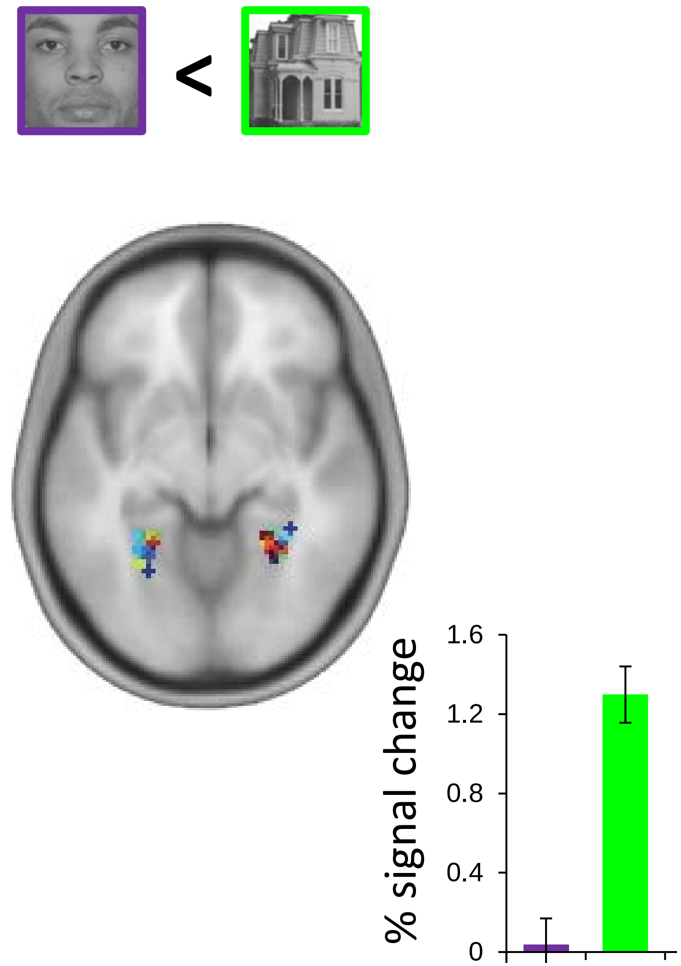
For example:
hippocampus

Step 3: select voxels
function (activation differences)

FFA



PPA

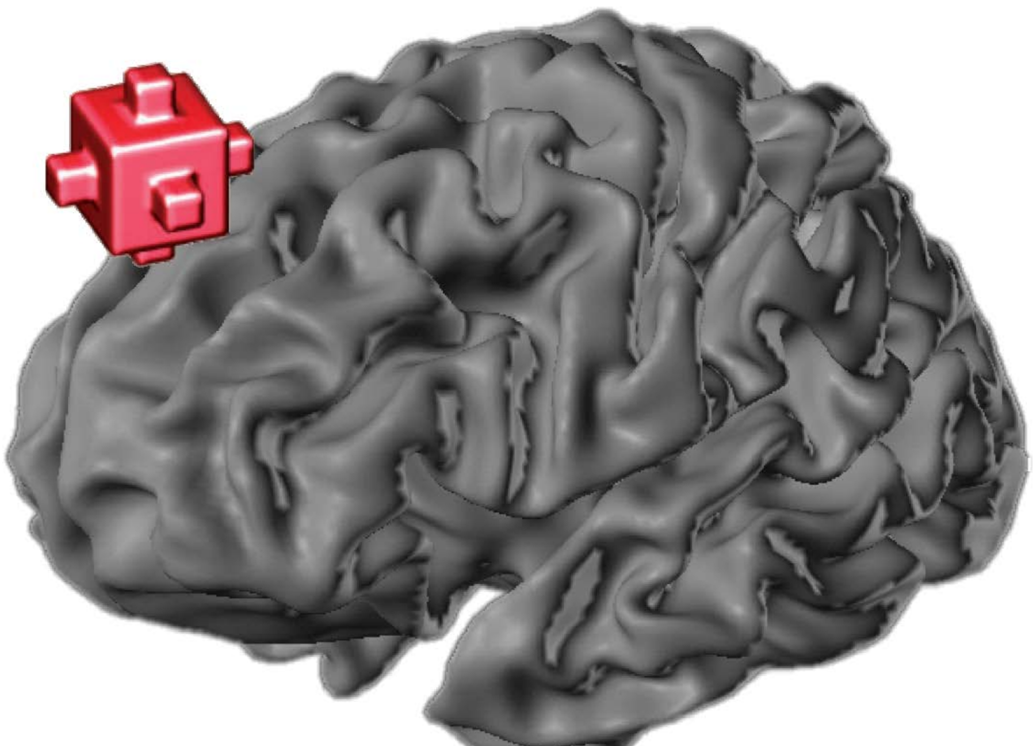


Step 3: select voxels

multivoxel searchlight

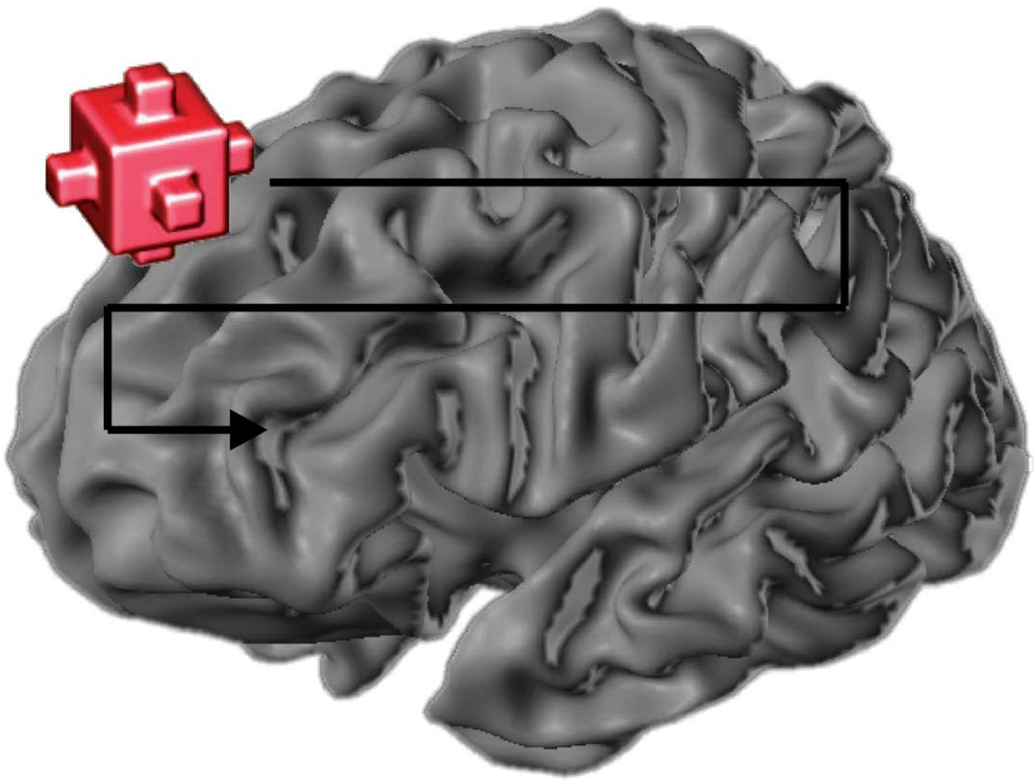
Step 3: select voxels

multivoxel searchlight



Step 3: select voxels

multivoxel searchlight



Kriegeskorte et al. 2006

Step 3: select voxels

How many voxels?

Depends on the expected spatial extent of effects.

Find the right balance:

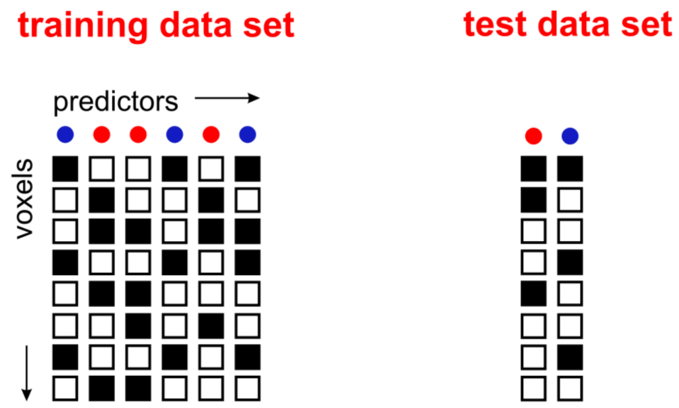
too few → risk of missing signal

too many → risk of overfitting (too noisy)

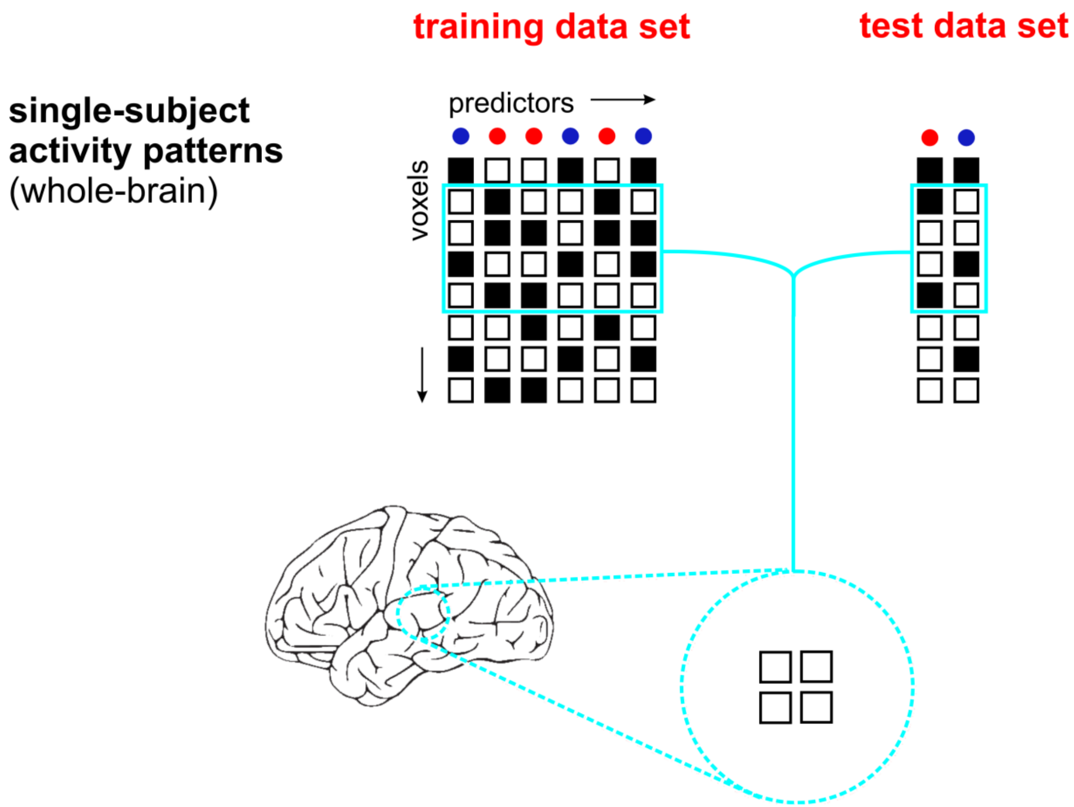
Common practice: select the same number of voxels in each subject.

Step 3: select voxels

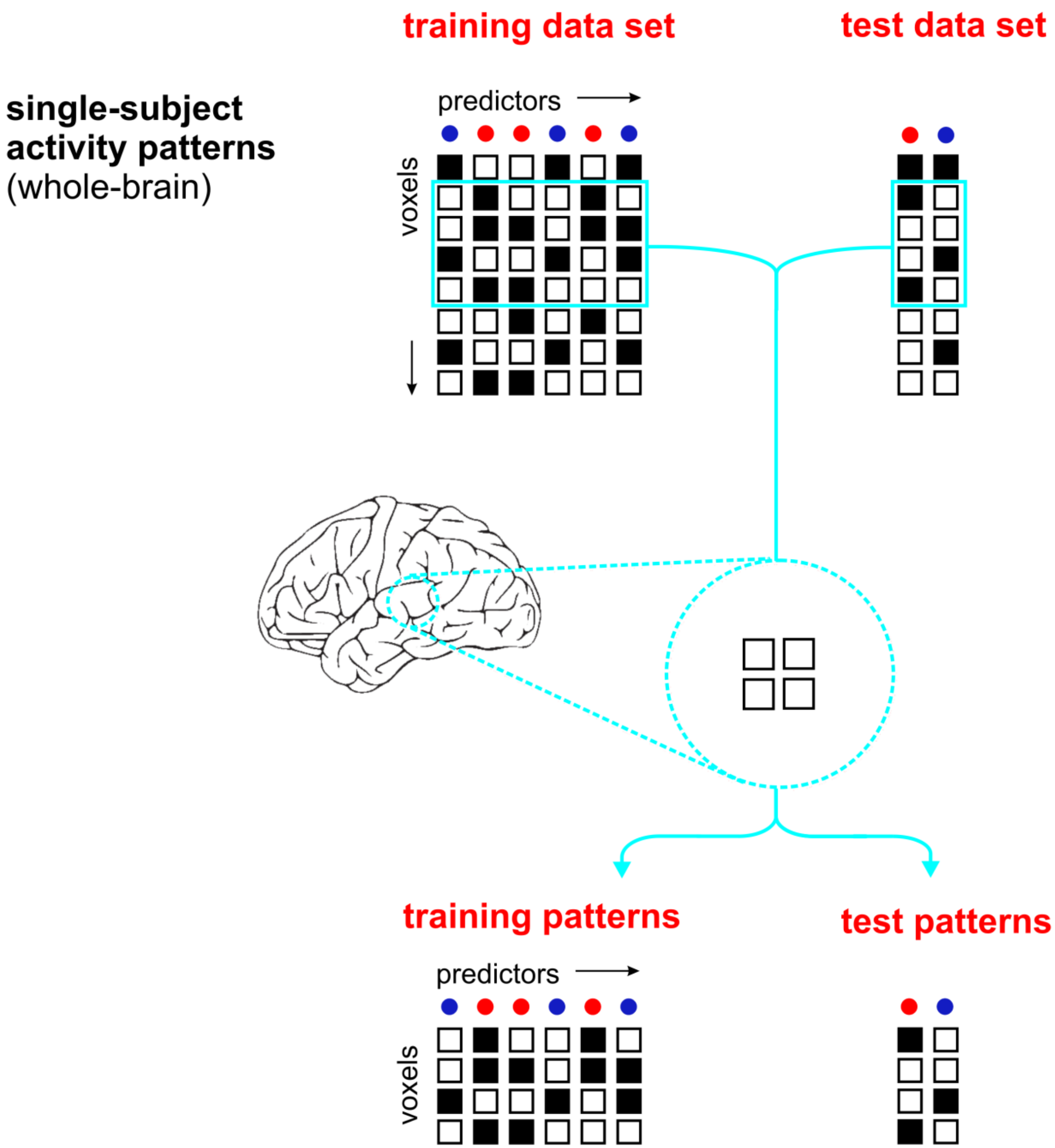
single-subject
activity patterns
(whole-brain)



Step 3: select voxels



Step 3: select voxels



Do it yourself: six steps

Step 1: split data and preprocess

Step 2: estimate single-subject activity patterns

Step 3: select voxels

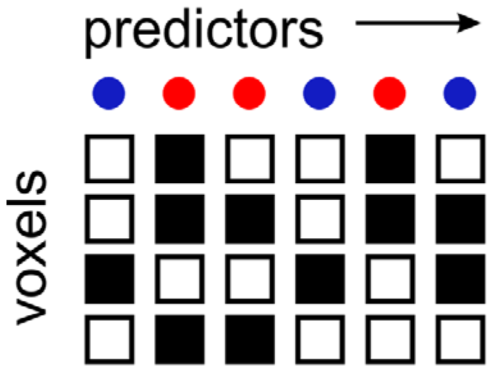
Step 4: train the classifier

Step 5: test the classifier

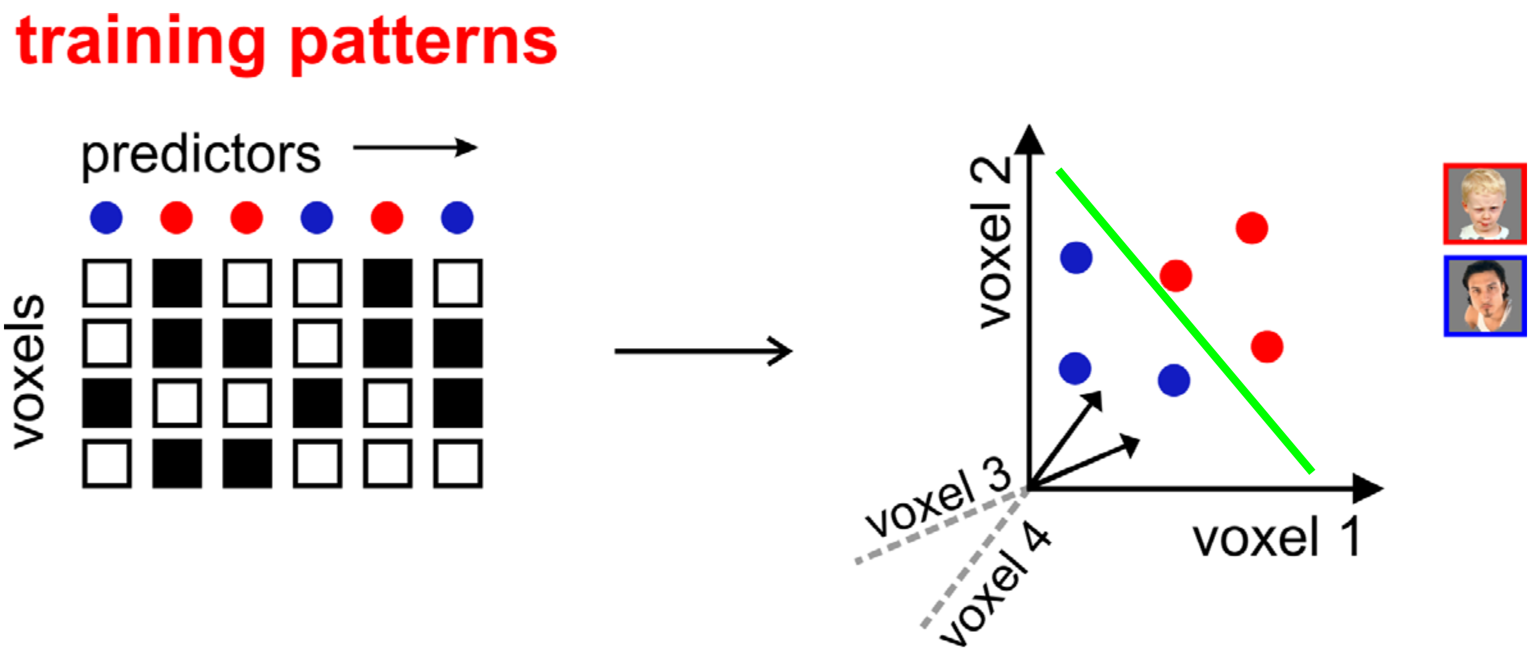
Step 6: statistical inference

Step 4: train the classifier

training patterns



Step 4: train the classifier



Do it yourself: six steps

Step 1: split data and preprocess

Step 2: estimate single-subject activity patterns

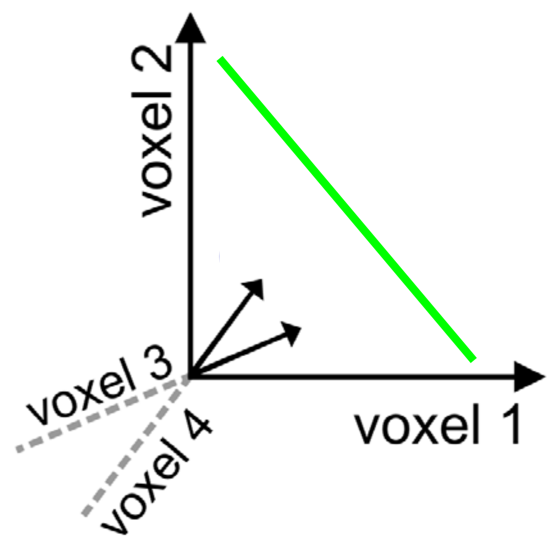
Step 3: select voxels

Step 4: train the classifier

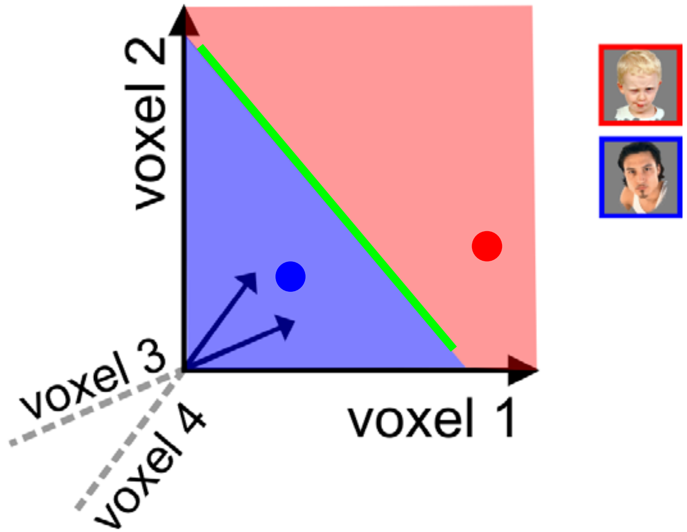
Step 5: test the classifier

Step 6: statistical inference

Step 5: test the classifier



Step 5: test the classifier



classification accuracy
for this fold = 100%

Cross-validation: generalise to….?

- different run (leave-run-out)
- different subject (leave-subject-out)
- different stimulus pair (leave-stimulus-pair-out)
- different block/trial within run (leave-block/trial-out)

Common procedure: use each run/subject etc as test data once.

For example: 4 runs → repeat cross validation 4 times (= 4-fold cross validation) → average accuracy across the 4 folds.

Do it yourself: six steps

Step 1: split data and preprocess

Step 2: estimate single-subject activity patterns

Step 3: select voxels

Step 4: train the classifier

Step 5: test the classifier

Step 6: statistical inference

Step 6: statistical inference

If number of subjects > 20:

Random-effects analysis across subjects using a standard one-sample right-sided t test.

H_0: $\mu = 50\%$

H_a: $\mu > 50\%$

Step 6: statistical inference

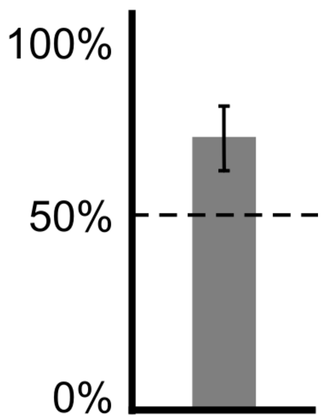
single-subject classification accuracy

error bars
= standard error
across *folds*

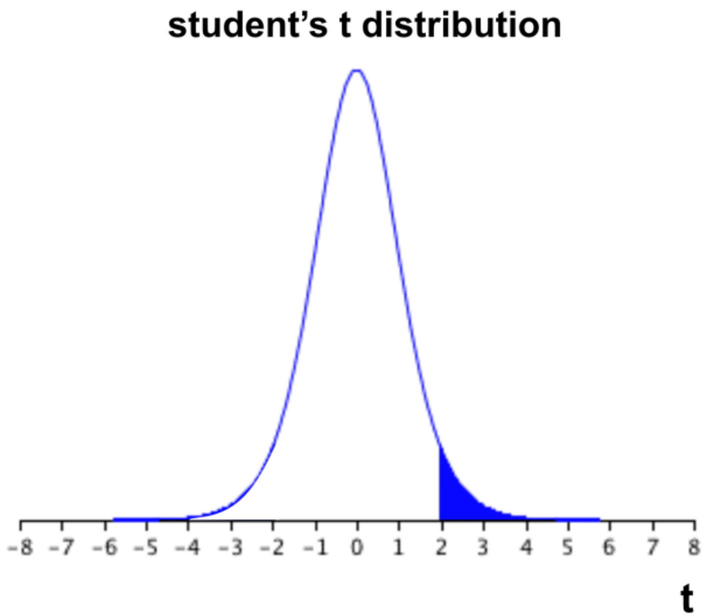
error bar
= standard error
across *subjects*

Step 6: statistical inference

subject-average classification accuracy



$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$



If the computed t value falls within the top 5% (blue) of the t distribution → reject H_0.

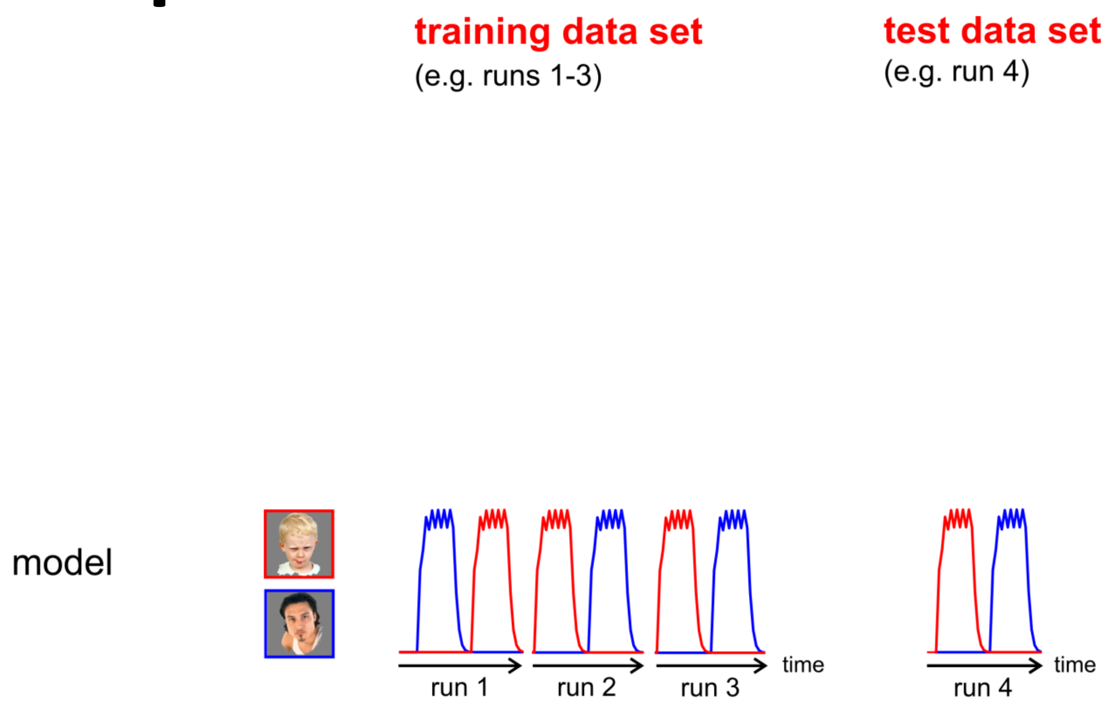
Step 6: statistical inference

If number of subjects <20:

We cannot assume a t distribution (central limit theorem does not apply)

→ use a permutation test: create a null distribution by randomly shuffling the condition labels during training.

Step 6: statistical inference

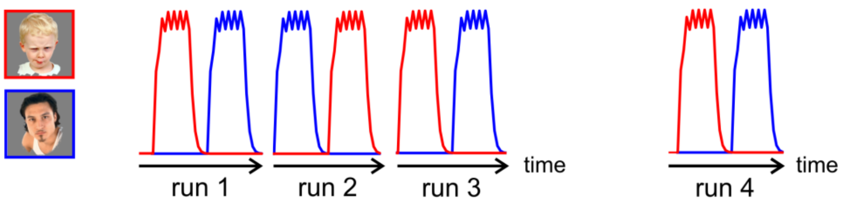


Step 6: statistical inference

training data set
(e.g. runs 1-3)

test data set
(e.g. run 4)

model



Remove the relationship between conditions and patterns.

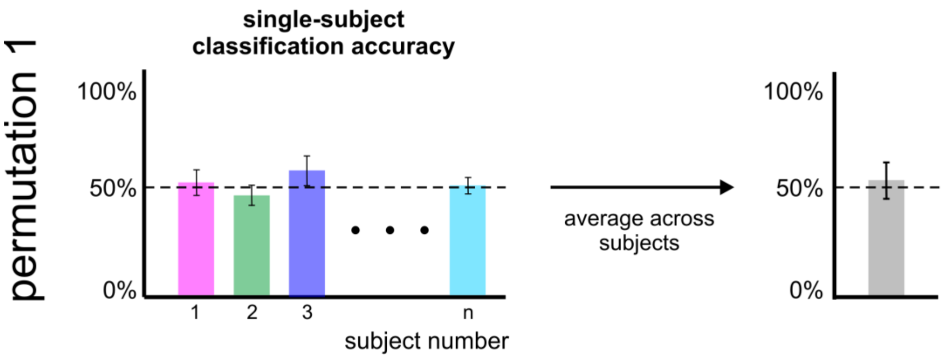
Step 6: statistical inference

Repeat step 2 – 5 after randomly reshuffling the condition labels.

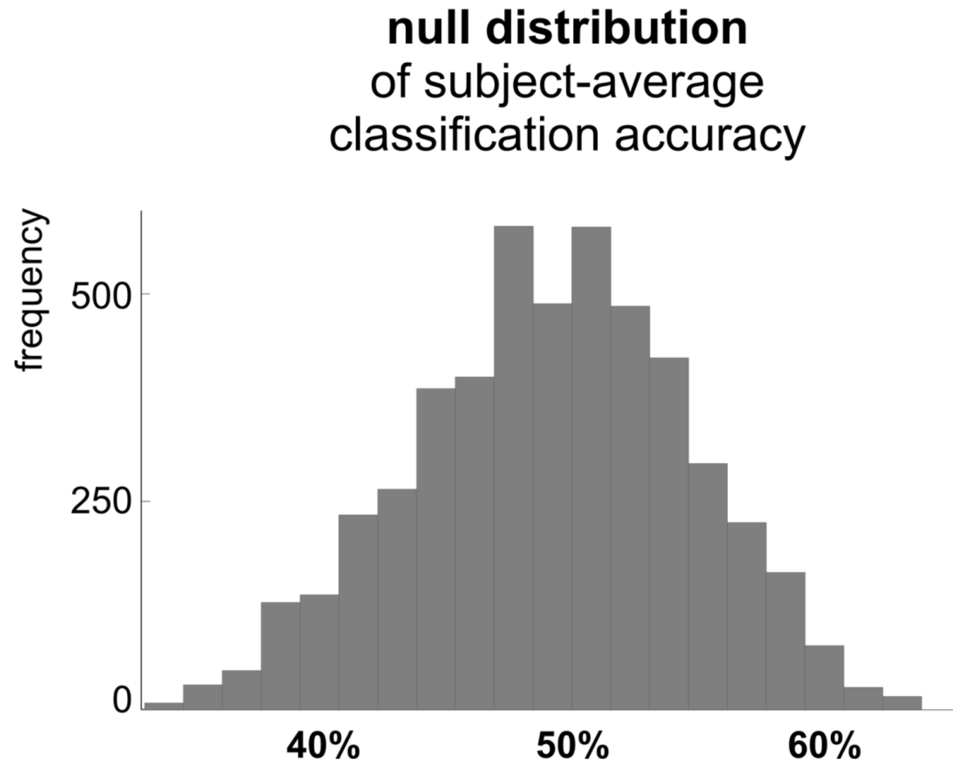
- step 2: estimate single-subject activity patterns
- step 3: select voxels
- step 4: train the classifier
- step 5: test the classifier

Do this many (e.g. 1000) times to create a null distribution.

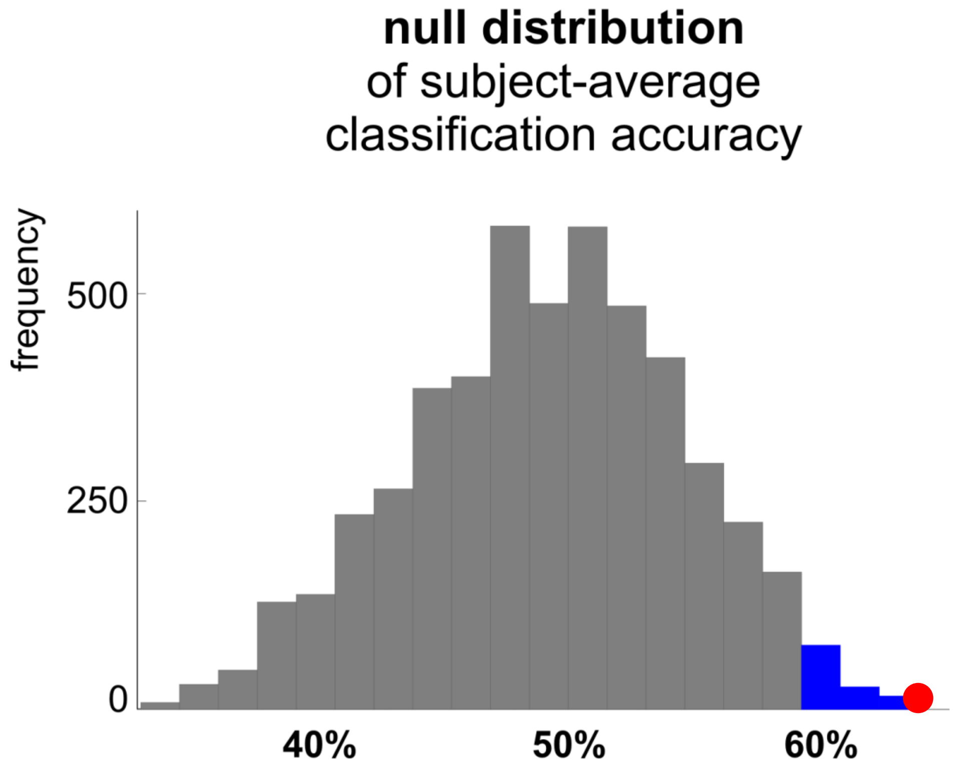
Step 6: statistical inference



Step 6: statistical inference



Step 6: statistical inference

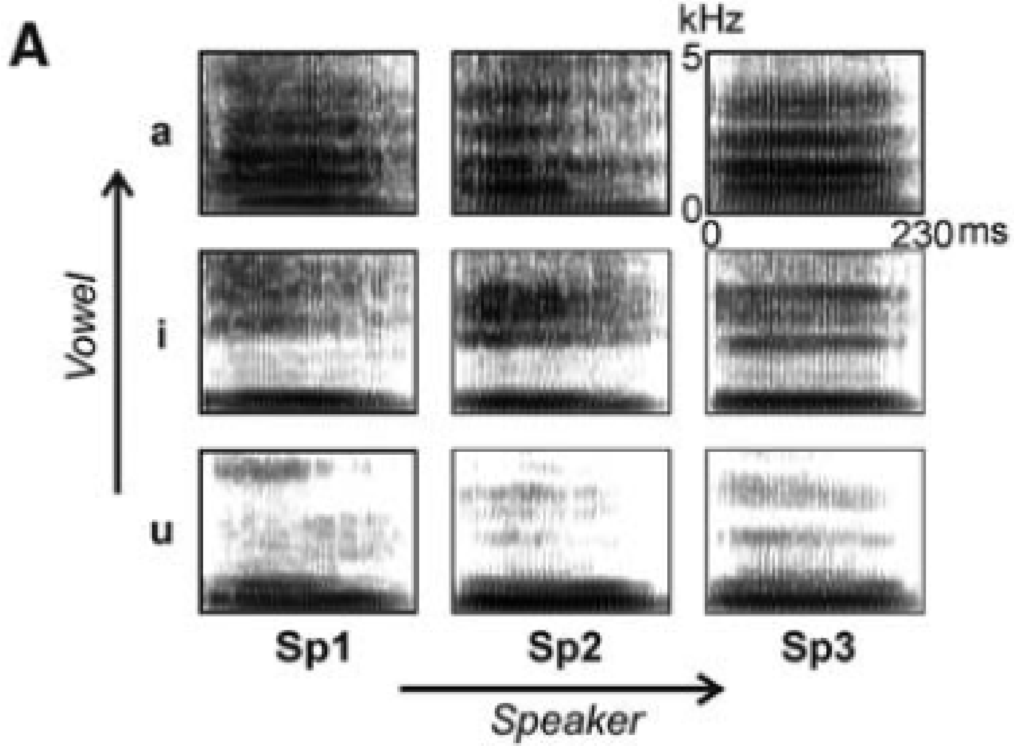


If the actual subject-average classification accuracy falls within the top 5% (blue) of the null distribution → reject H_0.

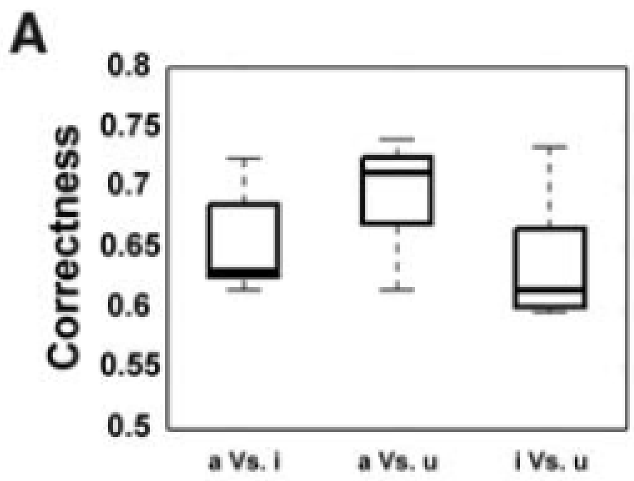
Applications: voice and speech

"Who" Is Saying "What"? Brain-Based Decoding of Human Voice and Speech

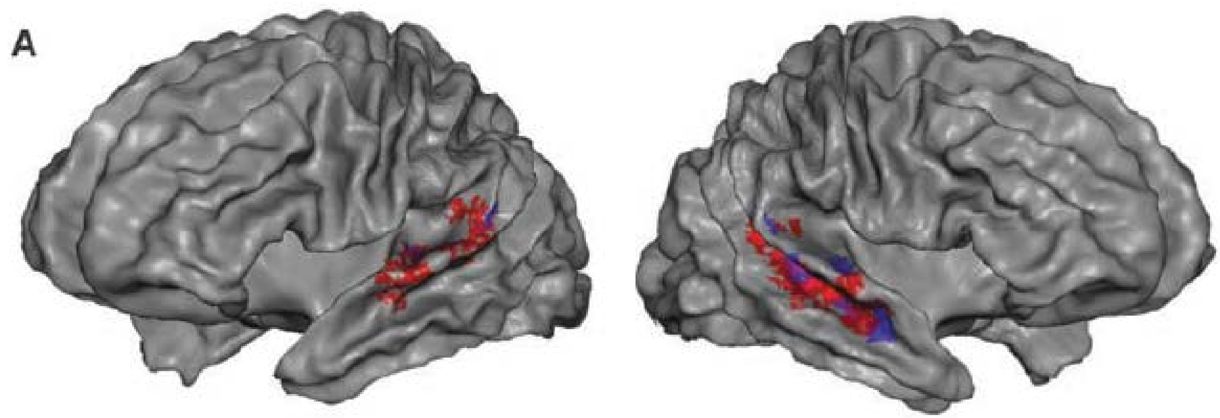
Elia Formisano,* Federico De Martino, Milene Bonte, Rainer Goebel



Applications: voice and speech



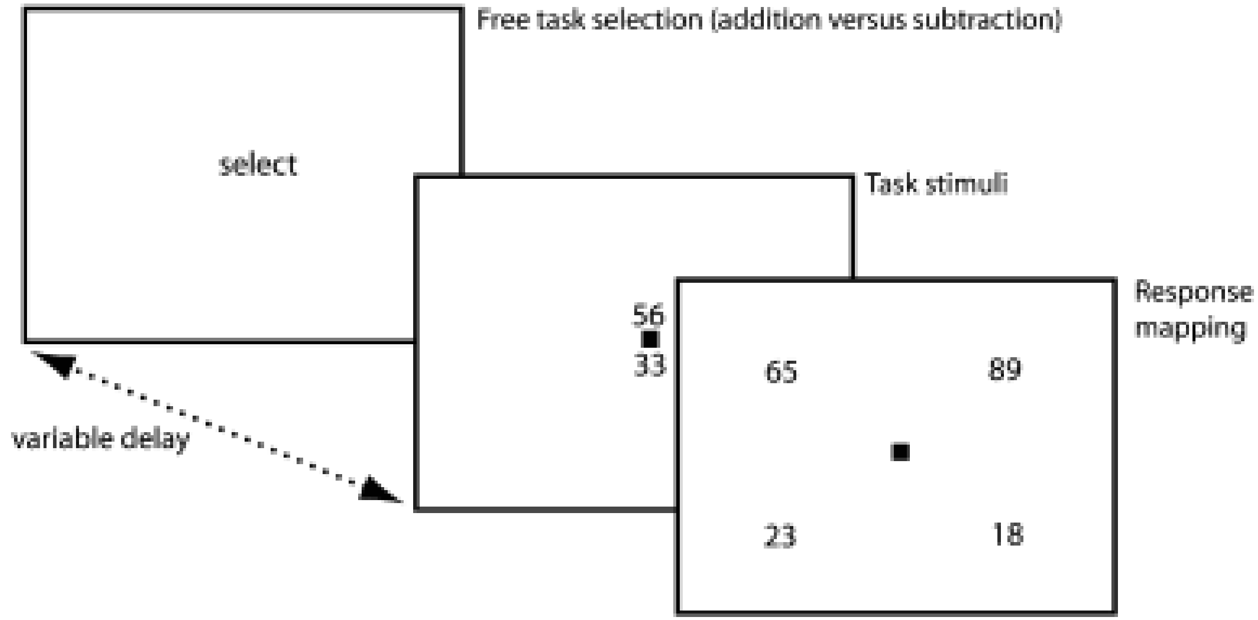
Applications: voice and speech



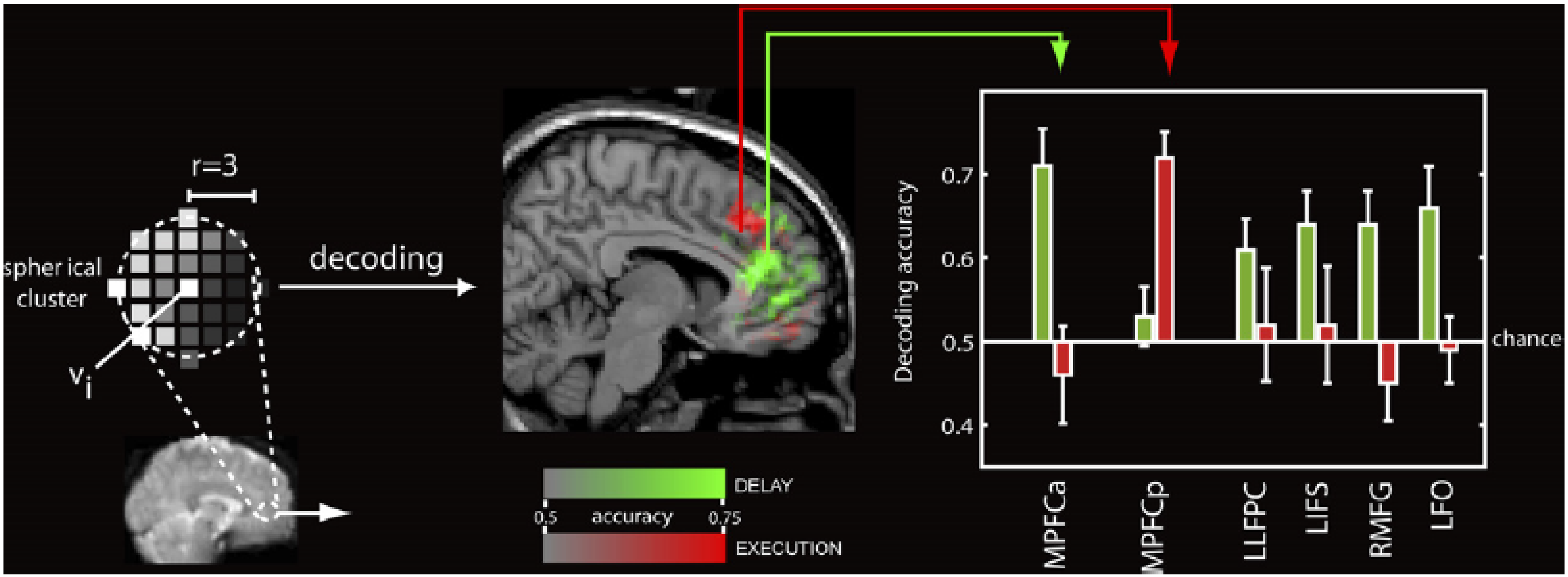
Applications: decision making

Reading Hidden Intentions in the Human Brain

John-Dylan Haynes,[1,2,3,4,5,*] Katsuyuki Sakai,[6] Geraint Rees,[4,5] Sam Gilbert,[4] Chris Frith,[5] and Richard E. Passingham[5,7]



Applications: decision making



Overview

- Why classification analysis?
- Linear classification: the basic idea
- Linear classification: different classifiers
- Do it yourself: six steps
  - step 1: split data and preprocess
  - step 2: estimate single-subject activity patterns
  - step 3: select voxels
  - step 4: train the classifier
  - step 5: test the classifier
  - step 6: statistical inference
- **Toolboxes**
- Literature

Toolboxes

- PRoNTo (SPM)

  http://www.mlnl.cs.ucl.ac.uk/pronto/

- LIBSVM

  http://www.csie.ntu.edu.tw/~cjlin/libsvm/

- PyMVPA

  http://www.pymvpa.org/

Overview

- Why classification analysis?
- Linear classification: the basic idea
- Linear classification: different classifiers
- Do it yourself: six steps
    - step 1: split data and preprocess
    - step 2: estimate single-subject activity patterns
    - step 3: select voxels
    - step 4: train the classifier
    - step 5: test the classifier
    - step 6: statistical inference
- Toolboxes
- **Literature**

Literature

Linear classification tutorials

Mur M et al. (2009) *Soc Cogn Affect Neurosci 4*: 101-109. [conceptual introduction]

Pereira F et al. (2009) *Neuroimage 45*(1 Suppl): S199-S209. [introduction]

Schreiber K, Krekelberg B (2013) *PLoS ONE 8*(7): e69328. [cautionary comments on statistical inference]

Kriegeskorte N et al. (2006) PNAS 103(10): 3863-3868. [multivariate searchlight]

Linear classification reviews

Norman KA et al. (2006) *Trends Cogn Sci 10*(9): 424-430.

Haynes JD, Rees G (2006) *Nat Rev Neurosci 7*: 523-534.

Linear classification: applications in neuroscience

Kamitani Y, Tong F (2005) *Nat Neurosci 8*(5): 679-685. [vision: classify orientations]

Formisano E et al. (2008*) Science 322*: 970-973. [voices: classify speakers & vowels]

Haynes JD et al. (2007) *Curr Biol 17*(4): 323-328. [cognitive control: task preparation]

Literature

Recursive feature elimination (RFE)

De Martino F et al. (2008) *Neuroimage 43*: 44-58.

Kernels

Jäkel F et al. (2009) *Trends Cogn Sci 13:* 381-388.

Which classifiers & preprocessing options are best?

Mourao-Miranda J et al. (2005) *Neuroimage 28*: 980-995. [SVM vs FLDA]

Kriegeskorte et al. (2009) *Nat Neurosci 12*(5): 535-540. [how to prevent selection bias]

Misaki M et al. (2010) *Neuroimage 53*: 103-118. [compares 6 different classifiers]

Garrido L et al. (2013) *Front Neurosci 7*(174): 1-4. [subtract the mean pattern?]

Relationships between classification (decoding), encoding, and RSA

Naselaris T et al. (2011) *Neuroimage 56*: 400-410.

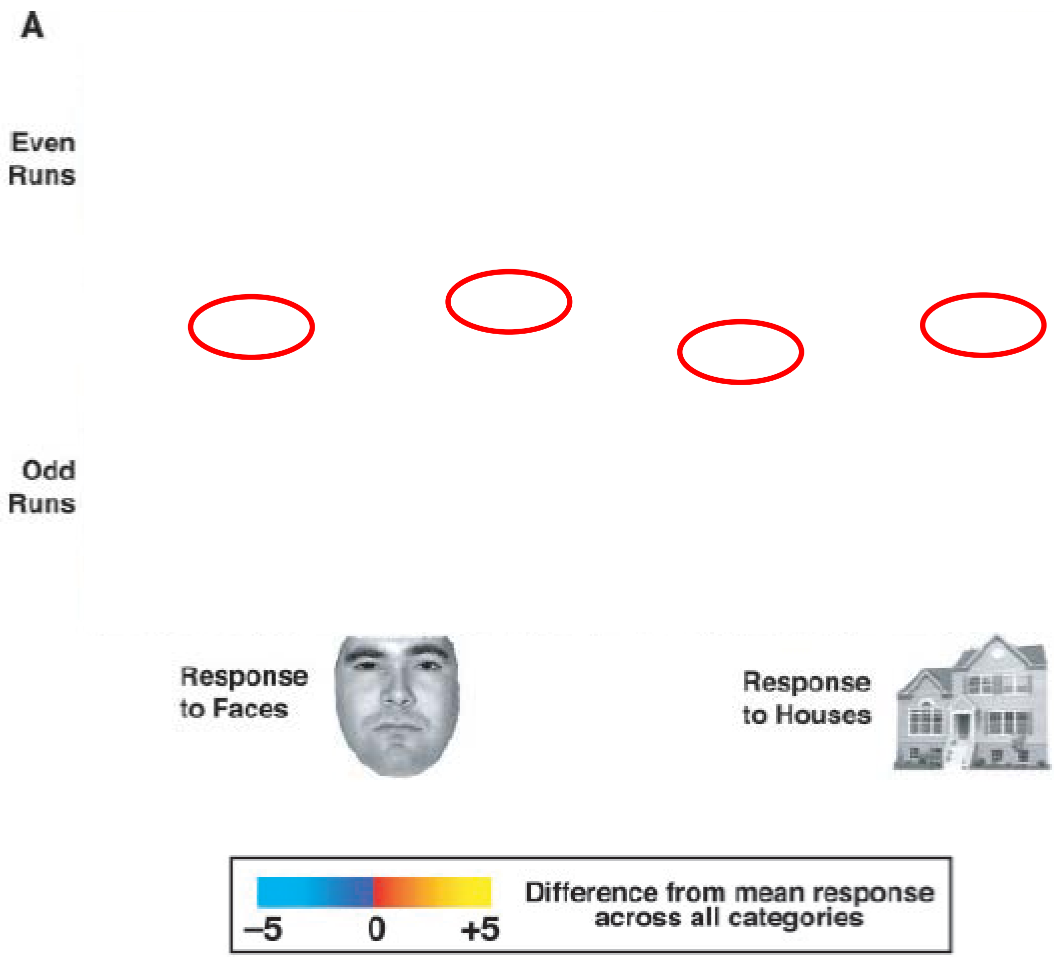
Kriegeskorte N (2011) *Neuroimage 56*: 411-421.

Example data set: Haxby et al. 2001

Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex

James V. Haxby,[1*] M. Ida Gobbini,[1,2] Maura L. Furey,[1,2] Alumit Ishai,[1] Jennifer L. Schouten,[1] Pietro Pietrini[3]

Example data set: Haxby et al. 2001



Example data set: Haxby et al. 2001

Can we discriminate faces and houses based on their whole-brain activity patterns?

Use a linear SVM in PRoNTo.

Set-up your laptop for the demo

To open matlab:

- Open terminal
- Type cdw
- Type matlab

Type in matlab:

- mkdir('/imaging/trainXXlinux/Workshop/Material/pronto/')
- addpath(genpath('/imaging/trainXXlinux/Workshop/Material/'))
- addpath(‘/hpc-software/matlab/r2009a/toolbox/stats/’)
- pronto