Session 5: Multivariate Analysis

Dr. Elena Galeano Weber

e.galeano-weber@dipf.de

Education and Human Development & Individualized Learning Laboratory, DIPF | Leibniz Institute for Research and Information in Education

IDeA Individual Development and Adaptive Education

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

DIPF Leibniz-Institut für Bildungsforschung und Bildungsinformation

Agenda Day 3

9:15 – 10:15
Session 5: Multivariate Analysis

10:15 – 11:45
Session 5: Hands-on Multivariate Analysis

11:45 – 12:45
Lunch Break

12:45 – 13:15
Session 6: Overview fMRI analysis pipeline

13:15 – 14:00
Session 6: Group work – “fMRI pitch”

14:00 – 14:30
Session 6: Group work – “fMRI pitch” presentation

14:30 – 15:00
Final countdown ;-)

Neural Decoding of Visual Imagery During Sleep

T. Horikawa,[1,2] M. Tamaki,[1]* Y. Miyawaki,[3,1]† Y. Kamitani[1,2]‡

Visual imagery during sleep has long been a topic of persistent speculation, but its private nature has hampered objective analysis. Here we present a neural decoding approach in which machine-learning models predict the contents of visual imagery during the sleep-onset period, given measured brain activity, by discovering links between human functional magnetic resonance imaging patterns and verbal reports with the assistance of lexical and image databases. Decoding models trained on stimulus-induced brain activity in visual cortical areas showed accurate classification, detection, and identification of contents. Our findings demonstrate that specific visual experience during sleep is represented by brain activity patterns shared by stimulus perception, providing a means to uncover subjective contents of dreaming using objective neural measurement.

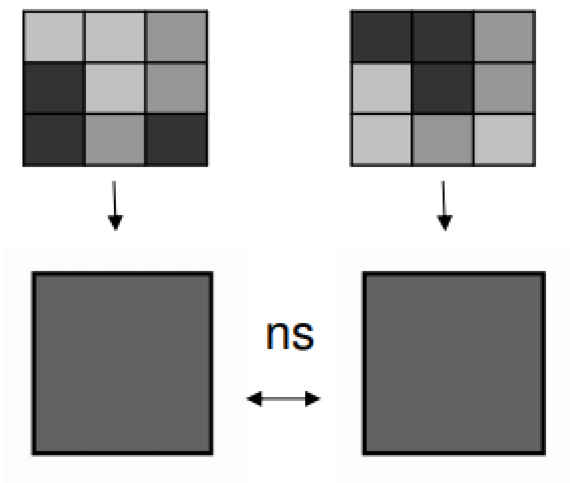
- https://www.zdf.de/play/dokus/die-grossen-fragen-mit-harald-lesch-100/die-grossen-fragen-haben-wir-eine-seele-mit-harald-lesch-doku-100

Why Multivariate in fMRI?

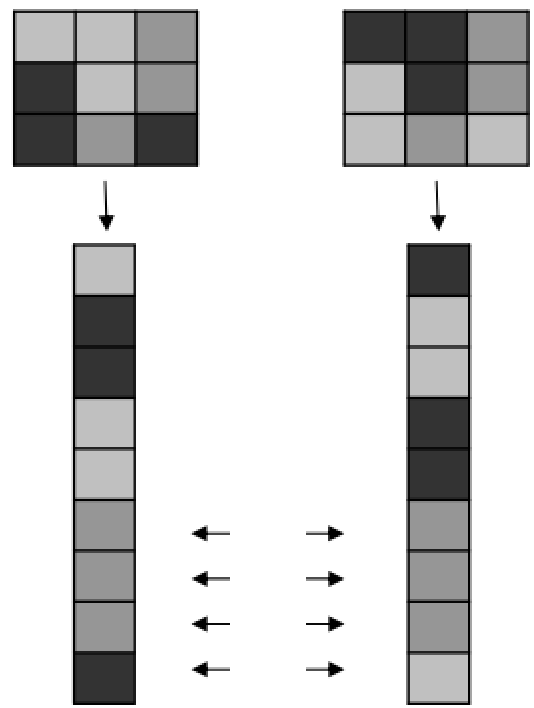
Feature	Univariate (GLM)	Multivariate (MVPA)
Focus	Voxel-level activation	Patterns across voxels
Sensitivity	Detect mean differences	Detect distributed info
Typical applications	Localization	Classification & decoding

Why Multivariate in fMRI?

Univariate:



Multivariate:



Why Multivariate in fMRI?

Fine grained localization: *where in the brain is it possible to distinguish between class A and class B?* (e.g., stimulus categories? stimulus identity?).

Representational content: what type of information is represented? (e.g., low level features vs. categorical information? detailed vs. coarse representations?

Temporal dynamics: From and until when is a given type of information available? (note: BOLD signal might not be the best measure for small time scales).

Why Multivariate in fMRI?

Classification.

Representational Similarity Analysis.

Core Concepts

- **Features**: *Each element inside an observation; voxel activation.*
- **Observations**: *Each one of the items inside a class that we will use in our analysis; trial-level beta map*
- **Class**: *Each "pattern" that we want find (we need at least two); experimental condition*
- **Classifying**: *Predict the type (i.e., class) of pattern that underlies a given dataset.*
- **Classifier:** *Model (e.g., SVM, CART)*
- **Decoding**: *From a given (brain) signal, figure out what caused the signal.*
- **Cross-validation:** Training/testing splits
- **Searchlight:** Local neighborhood decoding
- **MVPA** (multivoxel/multivariate pattern analysis): *Collection of analyses that use several sources of variance in the data to study it*.

Introduction to machine learning classifiers.

```
%% Practical example (human learning)
% If you see the following data sets that correspond to two conditions or classes,
% 1 and 2, can you see a pattern that distinguishes them?
array1 = [51; 53; 55; 57; 59; 61; 63; 65];
array2 = [32; 34; 36; 38; 40; 42; 44; 46];
labels1 = [1; 1; 1; 1; 1; 1; 1; 1];
labels2 = [2; 2; 2; 2; 2; 2; 2; 2];

% What is the best way of knowing whether you learn the pattern or not? Test your self:
test_observation(1)=34;

% Is test_observation a member of 1 or 2? Replace the "[]" with your answer.
your_prediction(1)=[];
```

Introduction to machine learning classifiers.

```
%% Can machines do that?
% Put all of the 'training' observations together
train_data=[array1; array2];
train_labels=[labels1;labels2];

% Train the model
model = fitcsvm(train_data, train_labels);

% Test its knowledge
model_prediction=predict(model, test_observation)
sprintf('The model thinks the observation "%d" belongs to class %d', ...
    test_observation(1), model_prediction)

% Compute performance
acc=model_prediction==test_label;
```

It very likely that the classifier agrees with your prediction. Do you know which pattern has it learned?

Introduction to machine learning classifiers.

Validation is the process of using fine-tuned tests to discover the pattern that the algorithm has learned.

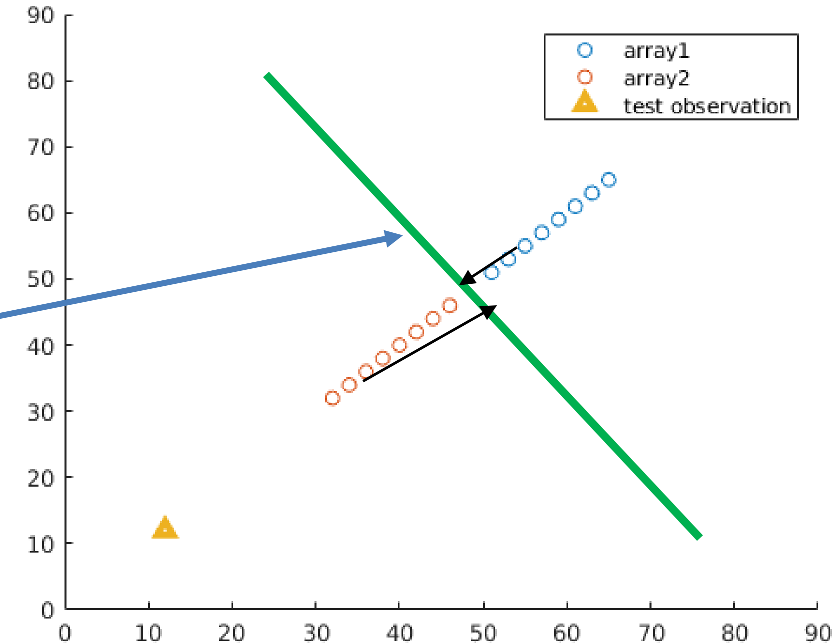
Let's do an exercise on independent validation.

```
%% Validation
% There is yet a better (more restrictive) way of assessing learning:
% Test on observations that are different from the study sets.
test_observation(2)=12;
test_label(2)=2;

% Is test_observation a member of 1 or 2? Replace the "[]" with your answer
your_prediction(2)=[];
```

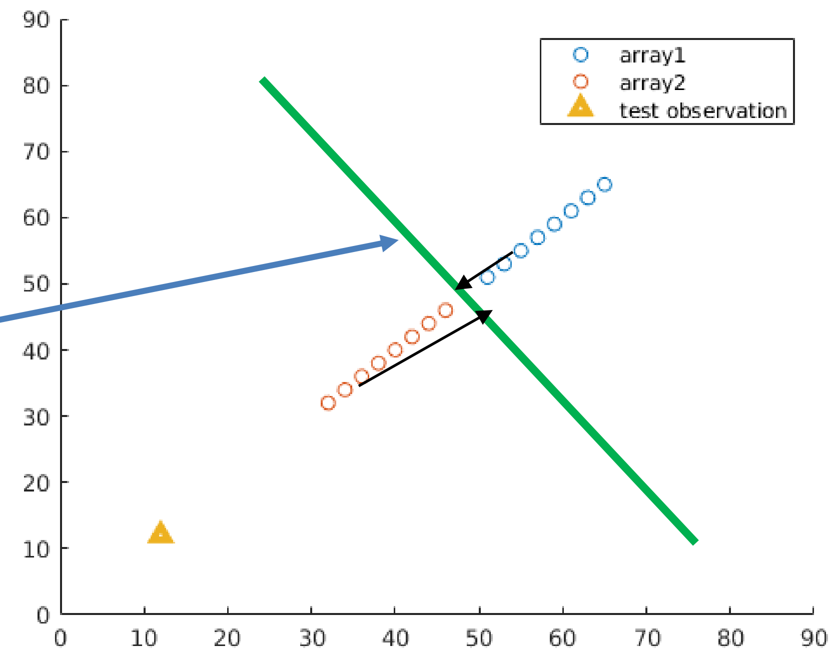
Introduction to machine learning classifiers.

Decision boundary



Introduction to machine learning classifiers.

Decision boundary

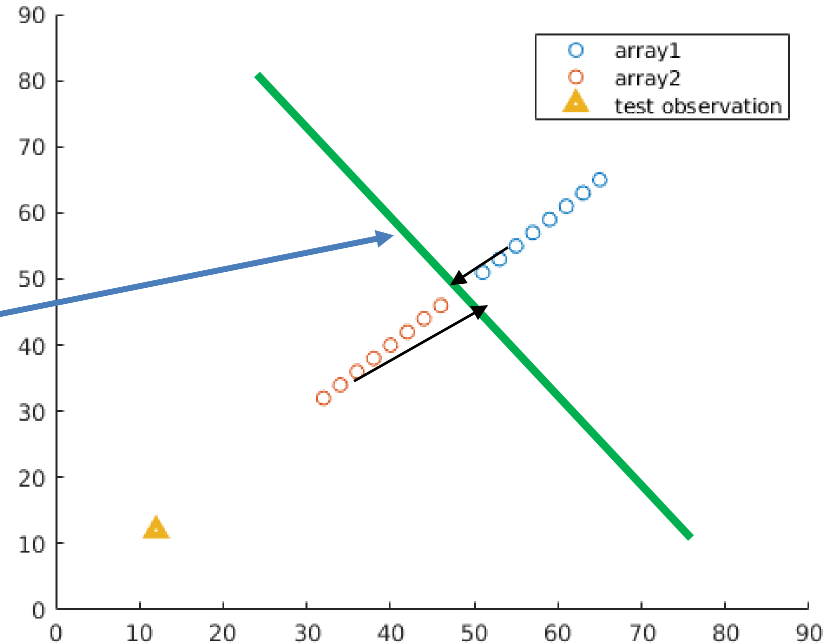


Why is independent validation necessary?

- Shows generalizability (avoids overfittings)
- Avoids overfitting

Introduction to machine learning classifiers.

Decision boundary



Problems:

- Limited data

Introduction. Validation on independent dataset .

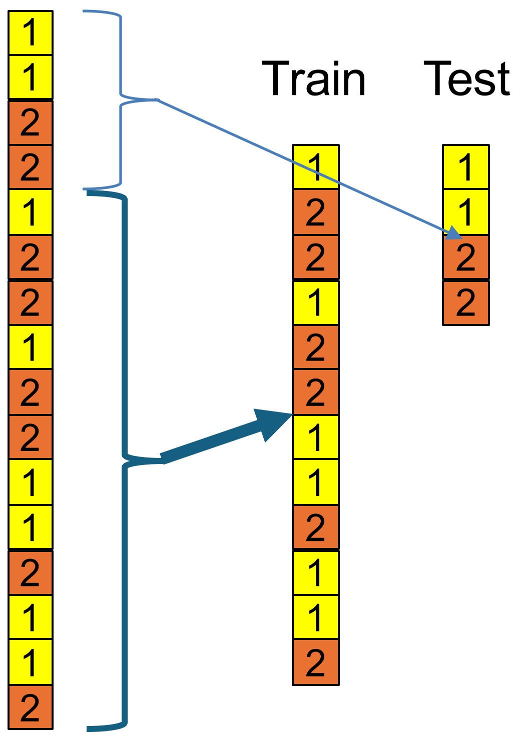
What can we do when we have limited data? Cross-validation.

Different cross-validation schemes allow for different (levels of) control or put more or less weight on the number of observations. Most common approaches:

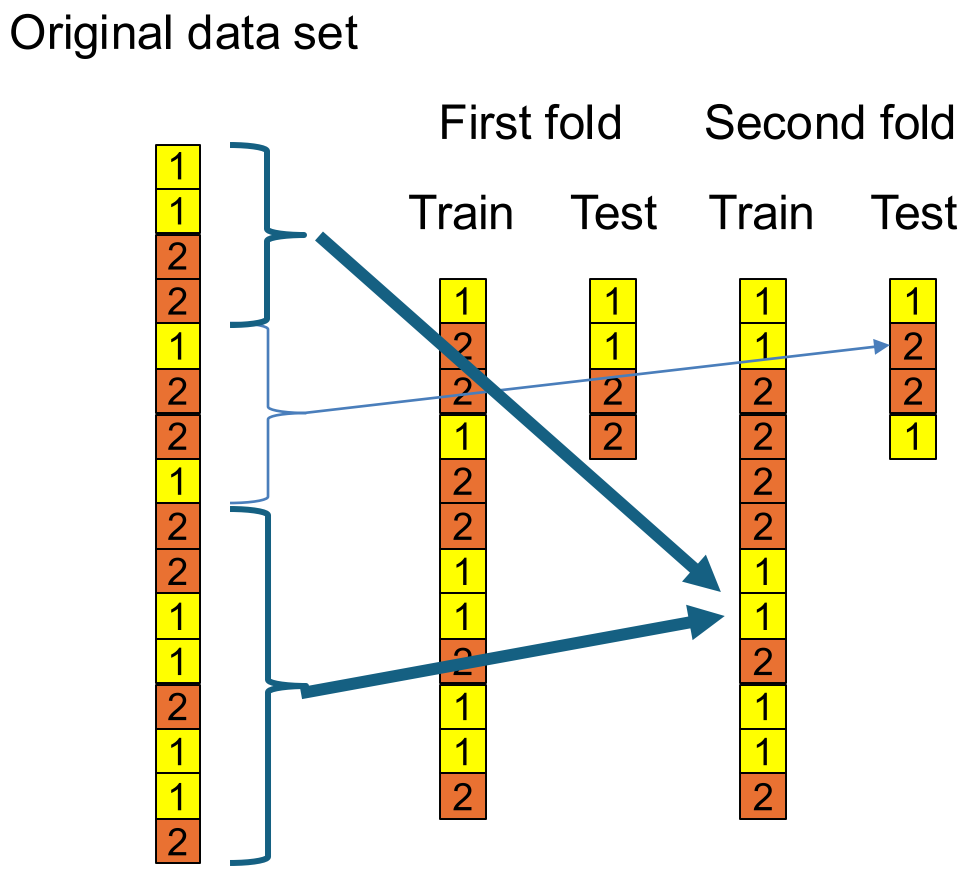
- Even/Odd observations (2 folds).
- Leave one (observation/chunk) out.

Introduction. Validation: Leave one (chunk) out .

Original data set

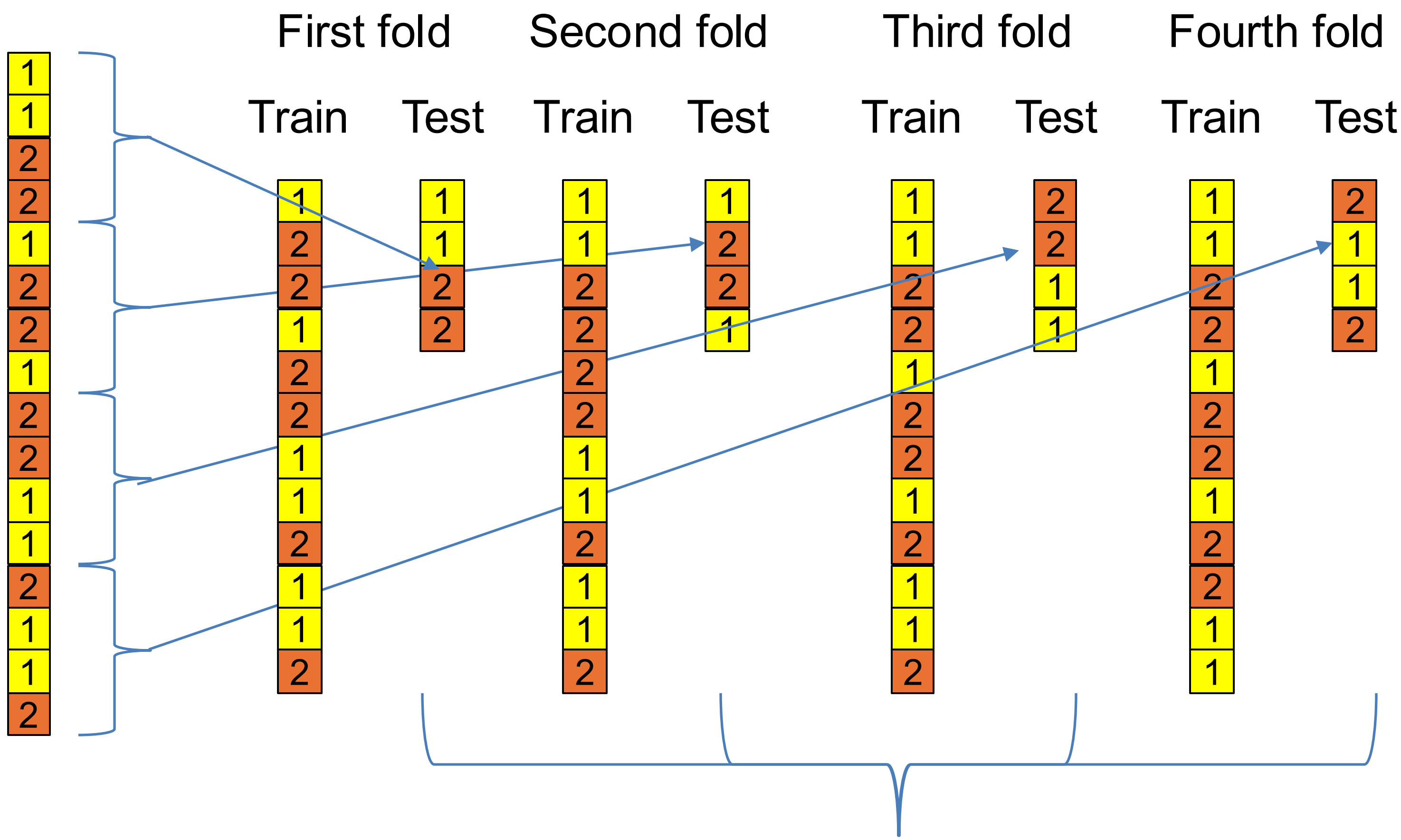


Introduction. Validation: Leave one (chunk) out .



Introduction. Validation: Leave one (chunk) out .

Original data set



Each observation has been validated on an independent dataset

Questions?

Classification in fMRI

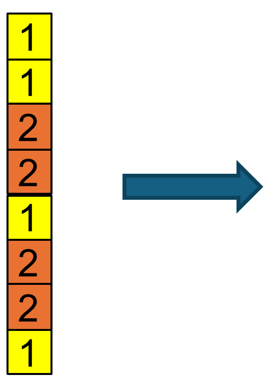
Classification in fMRI

Class: *Each "pattern" that we want find (we need at least two)*. In cognitive neuroscience classes are usually cognitive states (e.g., memory traces, task sets, representations).

Observations: *Each one of the items inside a class that we will use in our analysis*. Beta values (regressors in our GLM) of each condition.

Features: *Each element inside an observation*. Individual voxels (either in our ROI or in the entire brain).

8 observations x
1 feature



Classification in fMRI

Class: *Each "pattern" that we want find (we need at least two, e.g. experimental conditions)*. In cognitive neuroscience classes are usually cognitive states (e.g., memory traces, task sets, representations).

Observations: *Each one of the items inside a class that we will use in our analysis*. Beta values (regressors in our GLM) of each condition.

Features: *Each element inside an observation*. Individual voxels (either in our ROI or in the entire brain).

8 observations x
2 features

1	1
1	2
1.5	2
2	1.5
1	2
2	1.5
1.5	1
1	1

Classification in fMRI

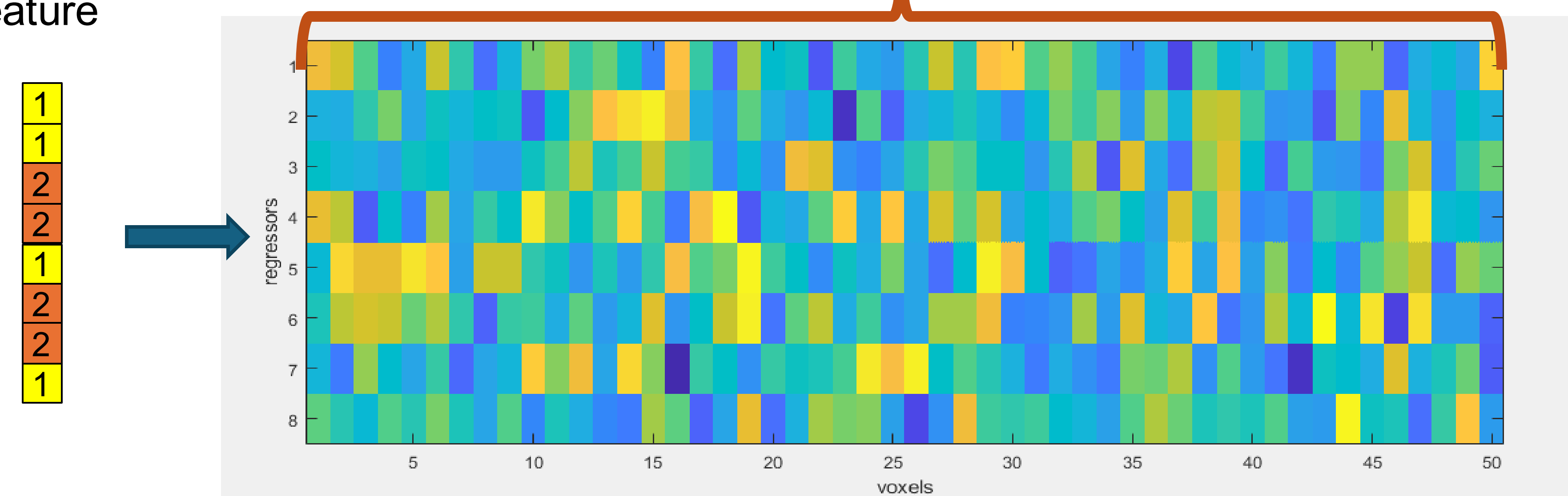
Class: *Each "pattern" that we want find (we need at least two)*. In cognitive neuroscience classes are usually cognitive states (e.g., memory traces, task sets, representations).

Observations: *Each one of the items inside a class that we will use in our analysis*. Beta values (regressors in our GLM) of each condition.

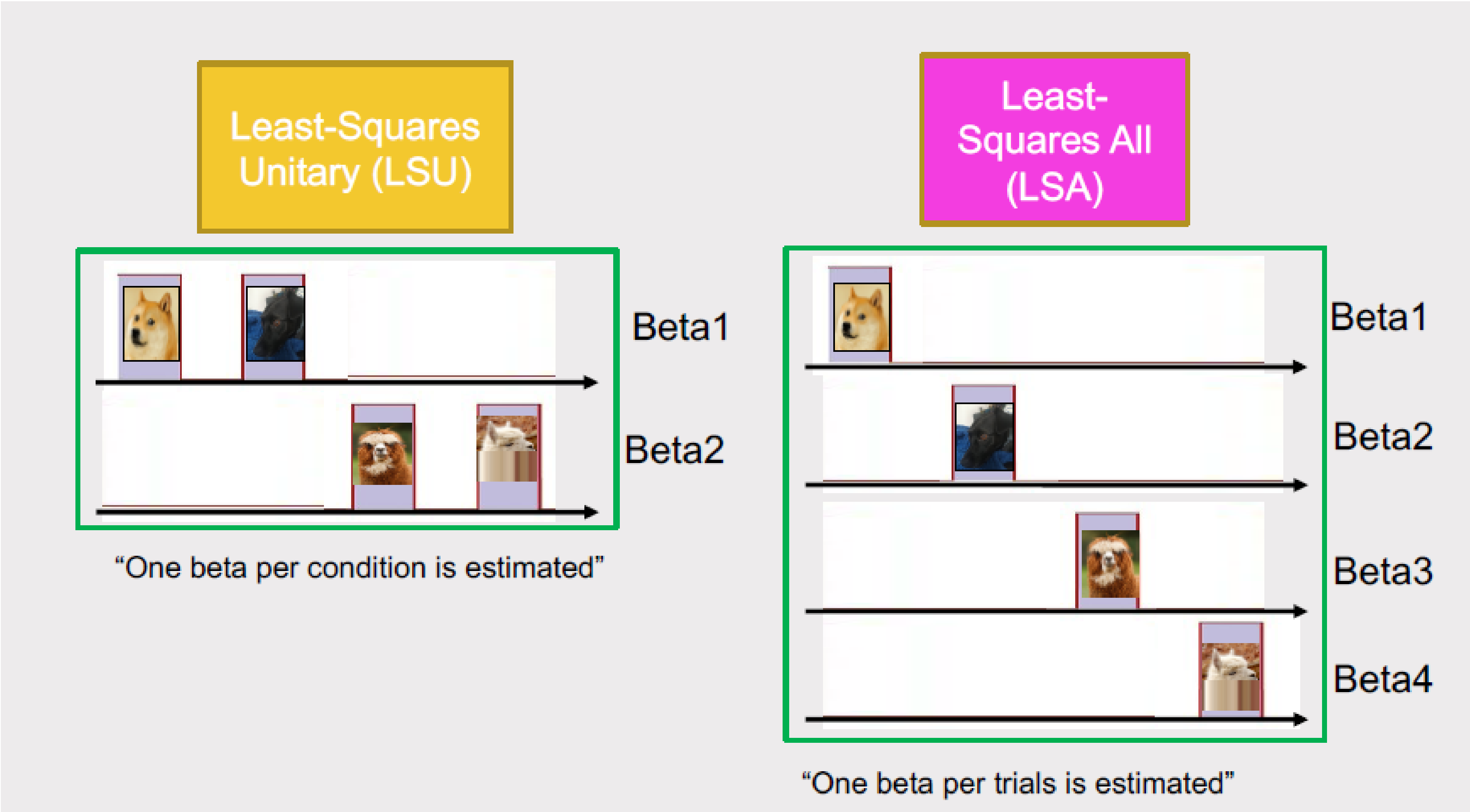
Features: *Each element inside an observation*. Individual voxels (either in our ROI or in the entire brain).

8 observations x 50 features

8 observations x
1 feature



Classification in fMRI.



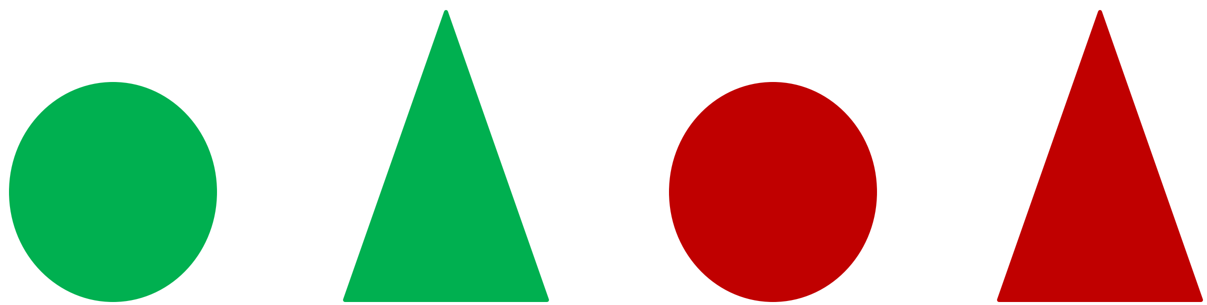
Rule of thumb:
"The number of observations must be bigger than the number of features"

Classification in fMRI. Example.

Simulated experiment.

Participants are exposed to circles and triangles in two different colors. They need to indicate the color of the shape by pressing one of two buttons.

Shapes are shown sequentially for 2 seconds and there’s a variable interval between shapes that ranges between 3s-15s.

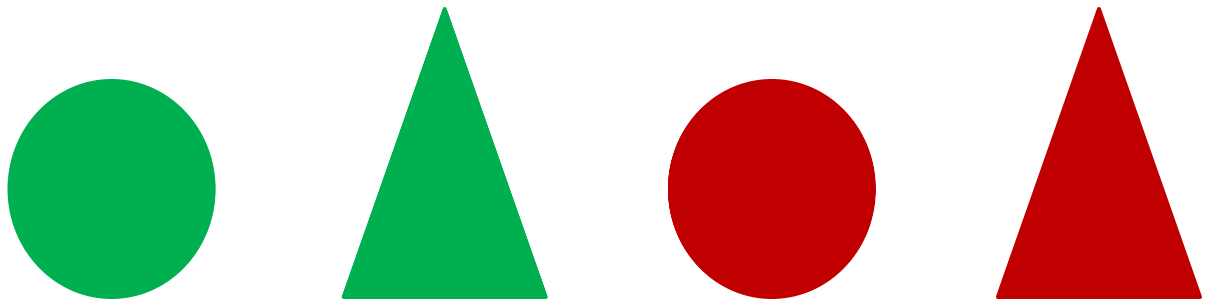


Classification in fMRI. Example.

Simulated experiment.

Participants are exposed to circles and triangles in two different colors. They need to indicate the color of the shape by pressing one of two buttons.

Shapes are shown sequentially for 2 seconds and there's a variable interval between shapes that ranges between 3s-15s.

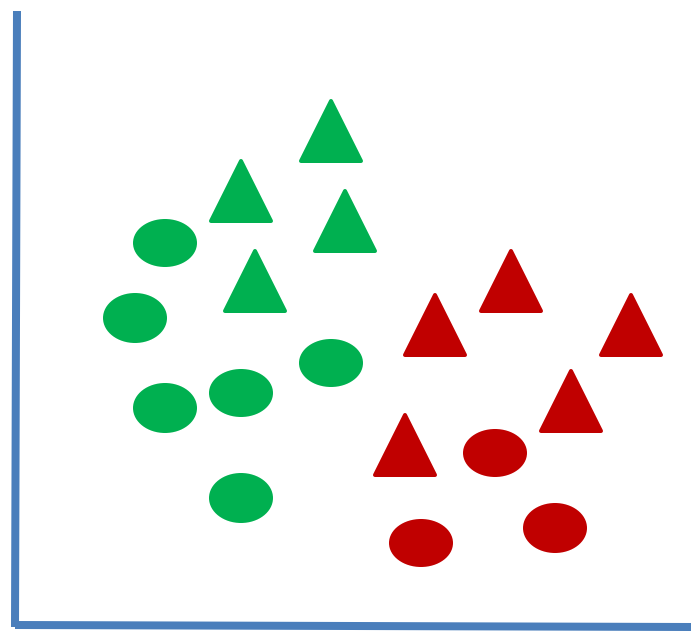


We have already estimated one beta per trial for a total of 16 trials on a full brain LSA GLM.

Classification in fMRI. Example.

Once we have our beta values. What do we do with them?

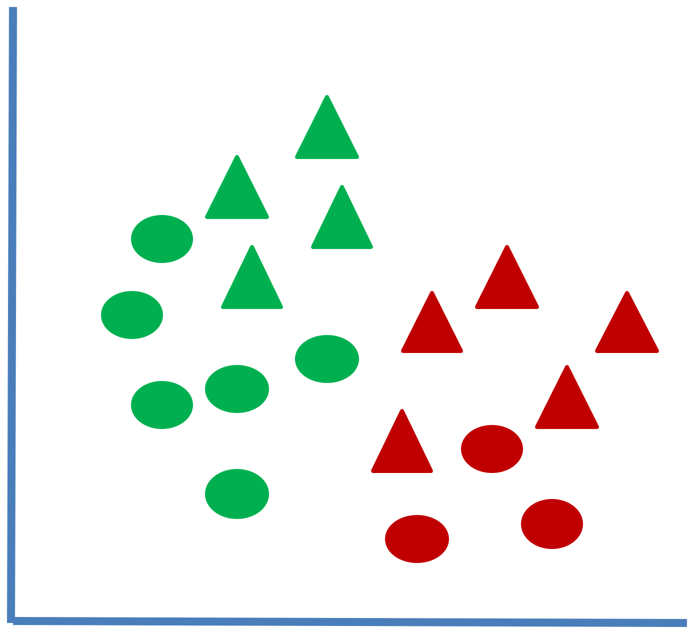
First decision: what do I want to classify?
Depending on the classes you select (i.e., the question that you ask the classifier), you can get different types of information.



Classification in fMRI. Example.

Once we have our beta values. What do we do with them?

First decision: what do I want to classify?
Depending on the classes you select (i.e., the question that you ask the classifier), you can get different types of information.

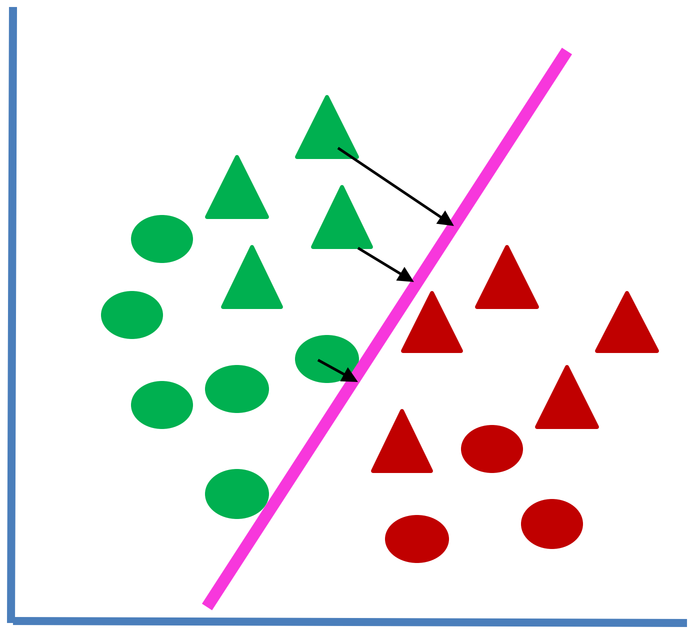


Red vs. green?

Classification in fMRI. Example.

Once we have our beta values. What do we do with them?

First decision: what do I want to classify?
Depending on the classes you select (i.e., the question that you ask the classifier), you can get different types of information.



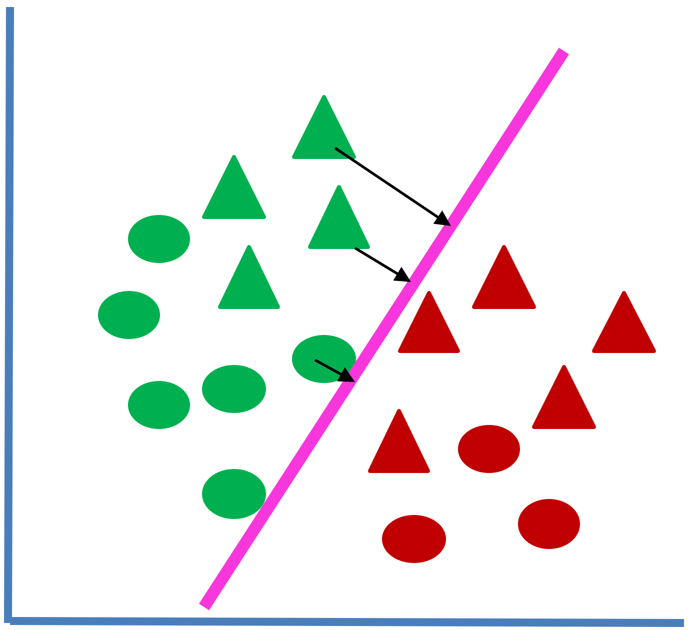
Red vs. green?

100%

Classification in fMRI. Example.

Once we have our beta values. What do we do with them?

First decision: what do I want to classify?
Depending on the classes you select (i.e., the question that you ask the classifier), you can get different types of information.

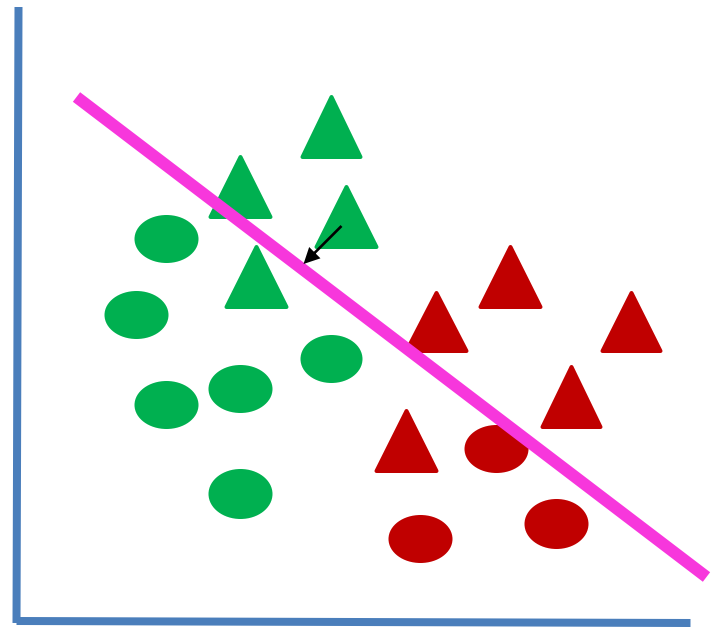


100%

Red vs. green?

or

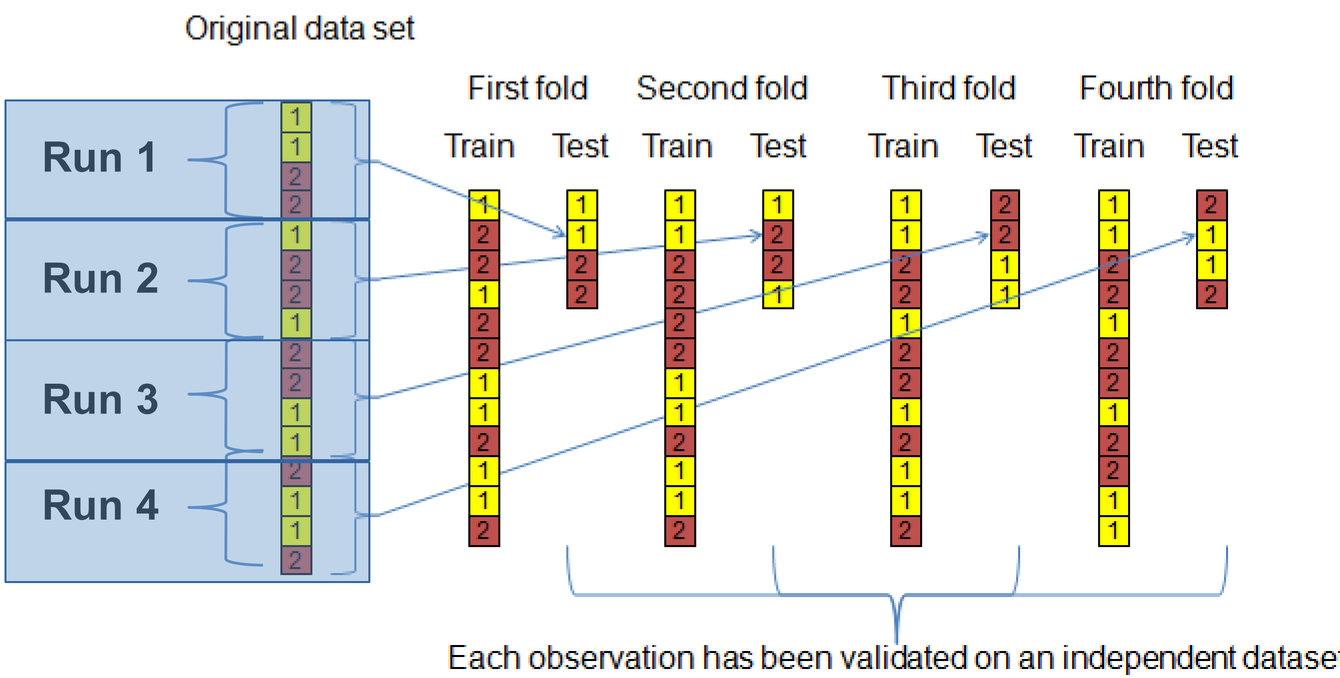
Circles vs. triangles?



?%

Classification in fMRI. Example.

Second decision: how will I validate my classifier?
The validation method can also add information (on top of controlling for biases; e.g., cross classification). In fMRI the standard method is **leave-one-run-out**: one scanning run is held for testing and training is performed with the remaining ones.

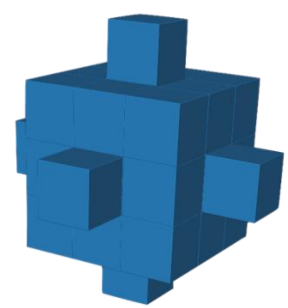


Classification in fMRI. Example.

Third decision: which voxels will I use (i.e., feature selection)? Two main approaches to consider:

- **Region of interest (ROI):** Only voxels within a given ROI are used in the analysis. ROIs can be defined either anatomically (e.g., right TPJ, left CA1) or functionally (e.g., only voxels that response more to faces than to houses). Functional localizers can be set up as an independent task (e.g., FFA localizer, retinotopy).

- **Searchlight**: a small cluster of voxels is selected for classification; the classification accuracy of that cluster is assigned to the central voxel in the cluster; then, an adjacent cluster is selected, and the process is repeated until the entire brain (or ROI) is covered.



Classification in fMRI. Example.

Fourth decision: how do I want to do it?

CoSMoMVPA. Oosterhof, Connolly, & Haxby, (2016). *Frontiers in neuroinformatics.* Multimodal environment. Works with matrices of data; requires the user to arrange datsets.
Matlab-based.

The Decoding Toolbox (TDT). Hebart, Görgen, & Haynes. (2015). *Frontiers in neuroinformatics.*
SPM-based. Works with beta maps (.nii) and SPM.mat from SPM; minimal user intervention for standard analysis.
Matlab-based.

PyMVPA. Hanke, Halchenko, Sederberg, Hanson, Haxby, & Pollmann, (2009). *Neuroinformatics*.
Python-based.

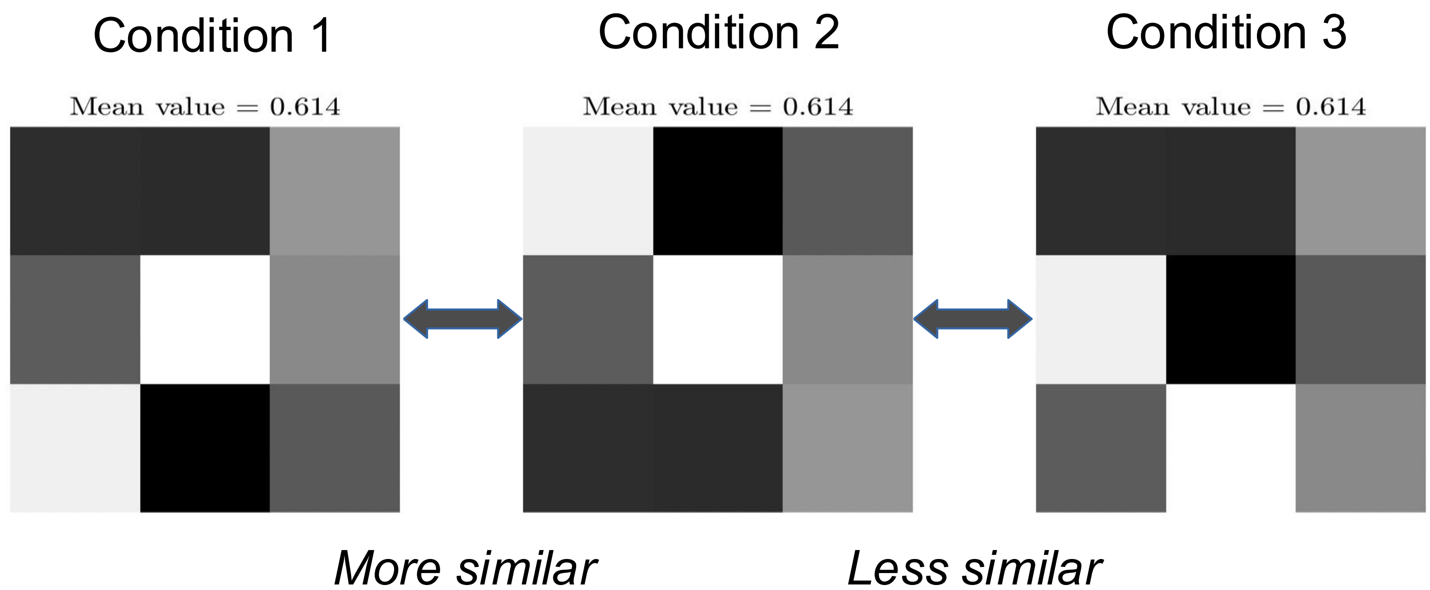
Questions?

Representational Similarity Analysis (RSA)

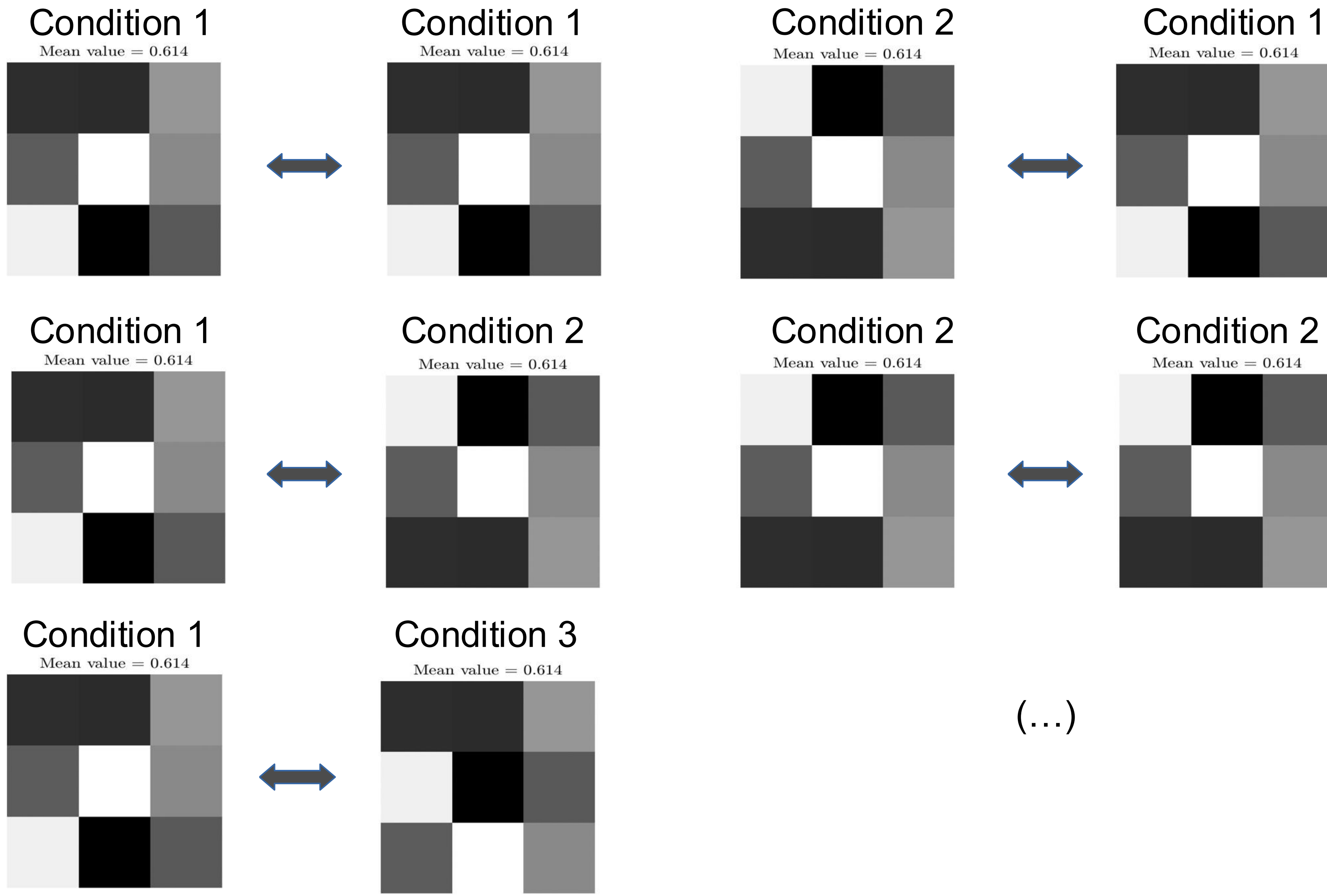
Kriegeskorte, Mur and Bandettini, 2008

Representational Similarity Analysis (RSA)

Using similarity between activity patterns to infer representational structure.



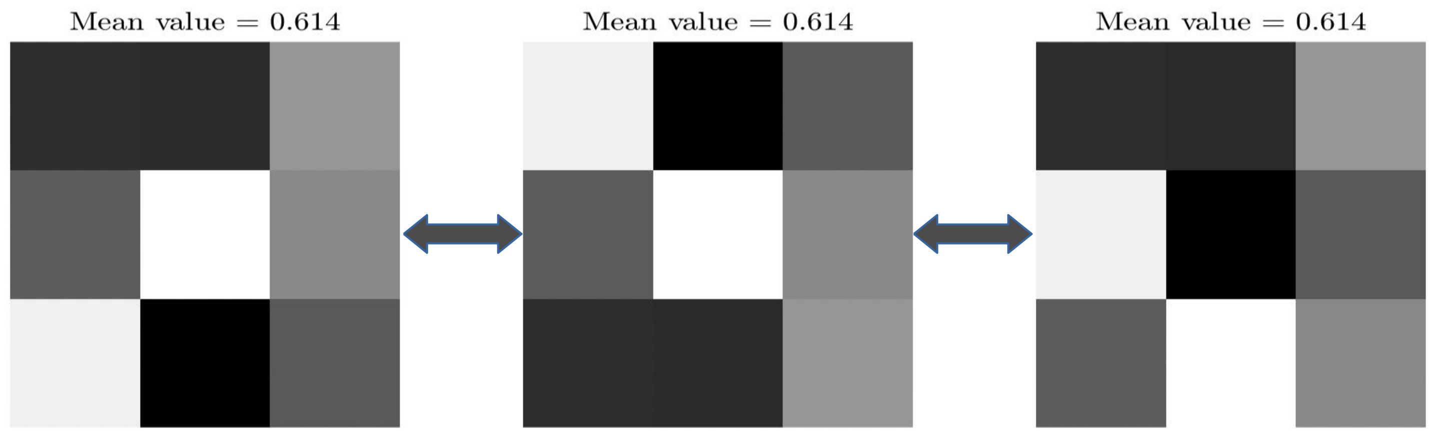
Representational Similarity Analysis (RSA)



Representational Similarity Analysis (RSA)

Using similarity between activity patterns to infer representational structure.

Condition 1 Condition 2 Condition 3



More similar
R = 0.29

Less similar
R = -.45

During condition 1 y 2, our ROI is representing the same *thing*;
during condition 3 is representing *something else*.

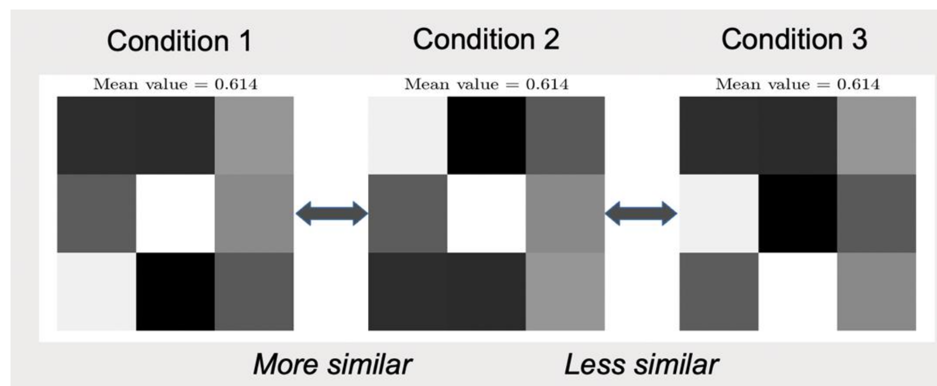
Representational Similarity Analysis (RSA)

Glossary of RSA jargon:

- **Similarity** = correlation / degree of overlap in activity patterns of a given brain region.

- **Dissimilarity** (a.k.a., representational distance): Degree of non-overlap in activity patterns of a given brain region. Quite often is 1 – correlation.

- **RDM**: Representational Dissimilarity Matrix. Pair-wise matrix of dis(similarity) values.



MATLAB Hands-On: Decoding and RSA with Simulated fMRI Data

→ Open Worksheet 5

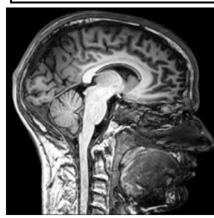
- Simulating fMRI-like voxel activation data with information about color and shape
- Performing classification to decode color or shape using LDA and SVM
- Extending classification analysis with feature importance and visualization
- Conducting RSA to compare neural representational structures with hypothesized models

Lunch Break

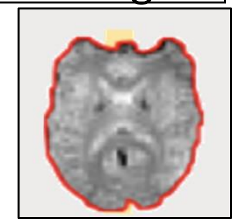
Overview: fMRI analysis pipeline

Magnetic properties of tissue

Anatomical images



Functional images

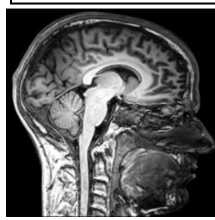


Overview: fMRI analysis pipeline

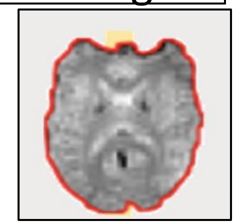
Magnetic properties of tissue

Anatomical images

Functional images



Preprocessing

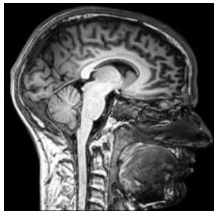


Overview: fMRI analysis pipeline

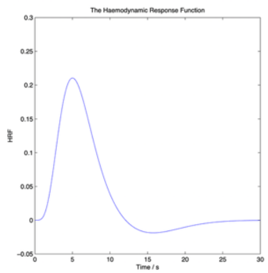
Magnetic properties of tissue

Anatomical images

Functional images



Preprocessing



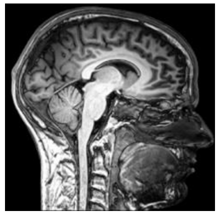
GLM

Beta maps

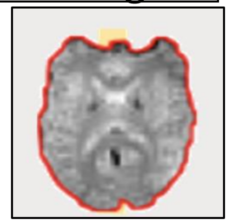
Overview: fMRI analysis pipeline

Magnetic properties of tissue

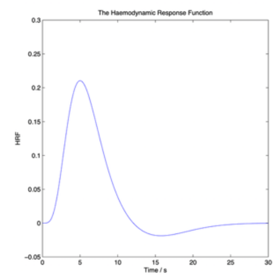
Anatomical images



Functional images



Preprocessing



GLM

Beta maps

Univariate

Contrasts

Group level

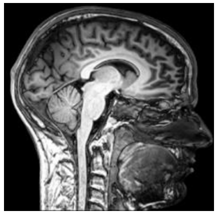
Activation maps

Overview: fMRI analysis pipeline

Magnetic properties of tissue

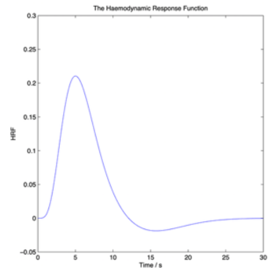
Anatomical images

Functional images



Preprocessing

GLM



Beta maps

Univariate

Multivariate

Contrasts

Classification

RSA

Group level

Activation maps

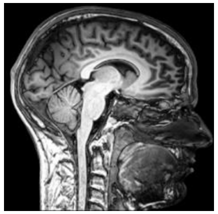
Classification accuracy

Similarity measures

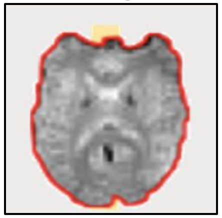
Overview: fMRI analysis pipeline

Magnetic properties of tissue

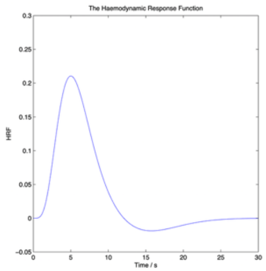
Anatomical images



Functional images



Preprocessing



Group-Level Interpretation

GLM

Beta maps

Univariate

Contrasts

Group level

Activation maps

Multivariate

Classification

Classification accuracy

RSA

Similarity measures

Final Group Task: "Pitch Your fMRI Study to a Funding Agency”

Apply everything you've learned in the workshop to **design and present** an fMRI study—from data acquisition to analysis strategy—targeted to address a meaningful cognitive neuroscience question. Your goal is to *convince a (fictitious) funding panel* (including myself ;-)) to fund your study.

The Scenario: You are a neuroscience research team preparing to submit a grant to the **Brain Imaging Research Council**. You have **10 minutes** to pitch your fMRI study idea. The panel wants to see that you:

- Understand fMRI methodology
- Can justify your design choices
- Know how to analyze and interpret fMRI data (both univariate and multivariate)
- Have a compelling scientific question

→ See for tips at the final worksheet