Intro to MVPA w/ CoSMoMVPA

Kyle Kurkela

kkurkela@bu.edu

Research Computing Services

Boston University

Resources

- Tutorial Slides and Materials
  - https://rcs.bu.edu/examples/imaging/tut_mvpa/
- *CoSMoMVPA* Website
  - https://cosmomvpa.org
- Questions? Problems? Help?
  - help@scc.bu.edu
  - kkurkela@bu.edu

Tutorial Feedback

Please open a tab in your web browser and have this page open:

http://scv.bu.edu/survey/tutorial_evaluation.html

Please fill it out at the end of todays tutorial!

Learning Objectives

- ❑ Understand basic MVPA concepts
- ❑ Understand how data are organized in CoSMoMVPA
- ❑ Be able to run a basic classification analysis
- ❑ Be able to run a basic searchlight analysis
- ❑ Be able to perform a permutation test

Outline

- Tutorial Setup
- MVPA Background
- CoSMoMVPA Data Model
- Classification and Cross-Validation
- Searchlights
- Statistical Inference
- Wrap Up

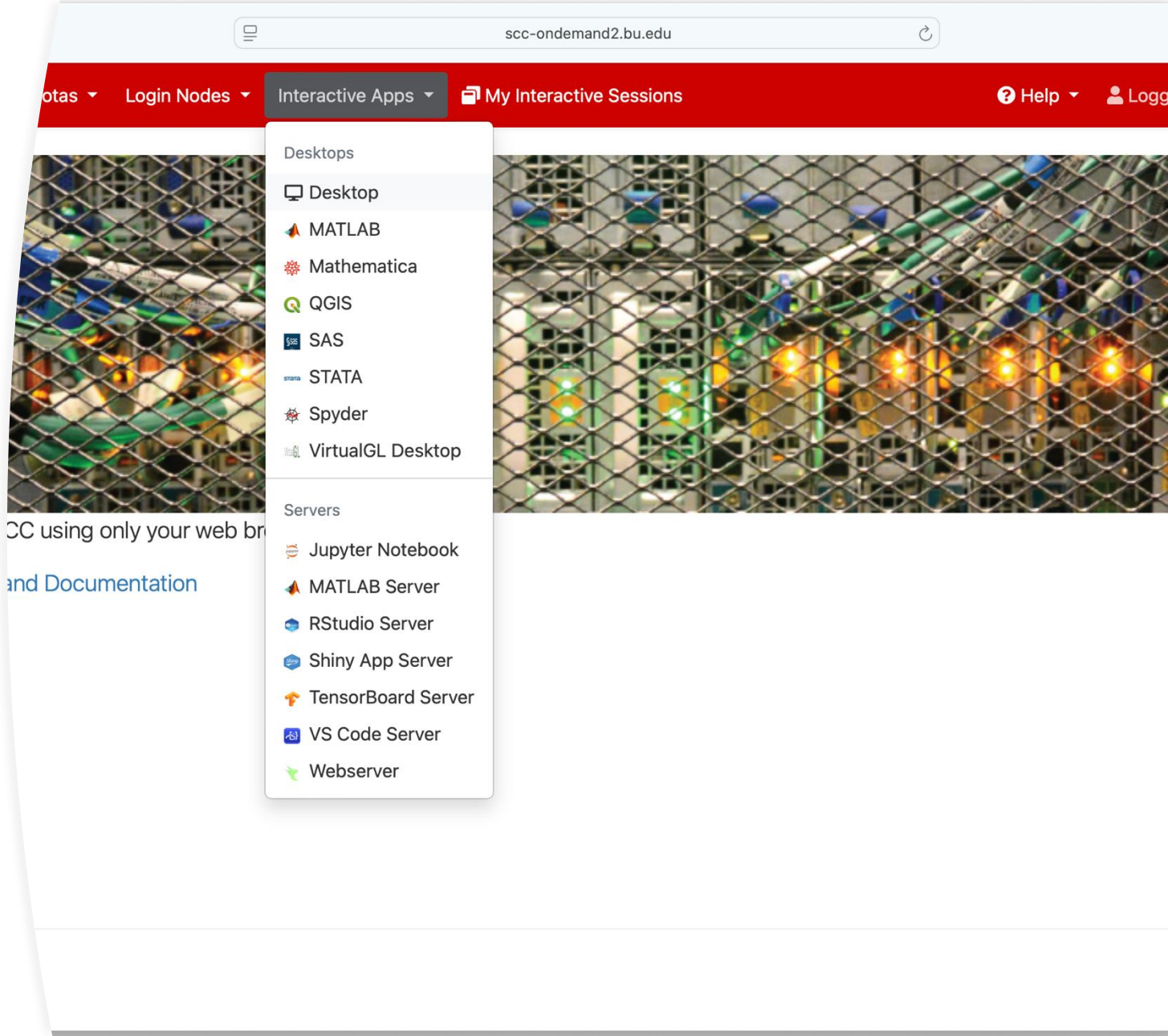
Tutorial Setup

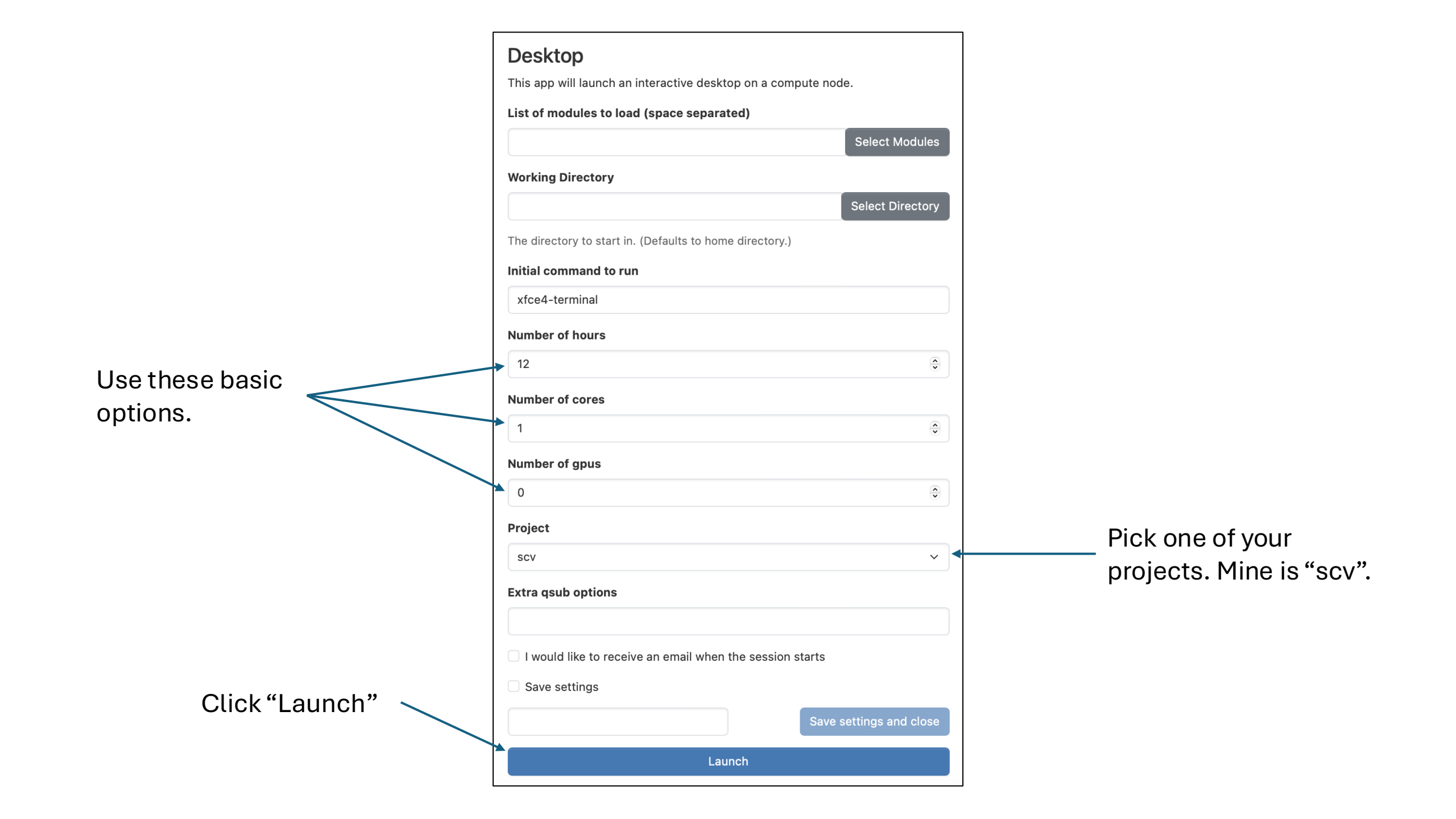
Open an OnDemand Session

1. Go to scc-ondemand.bu.edu
2. Go to Interactive Apps
3. Request a Desktop

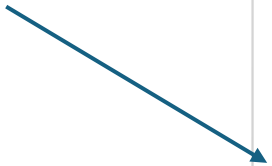
Tutorial Accounts

1. Go to scc-ondemand-tutorial.bu.edu
2. Username and Password will be sent in Zoom chat
3. Go to Interactive Apps
4. Request a Desktop




Desktop
This app will launch an interactive desktop on a compute node.
List of modules to load (space separated)
Select Modules
Working Directory
Select Directory
The directory to start in. (Defaults to home directory.)
Initial command to run
xfce4-terminal
Number of hours
12
Number of cores
1
Number of gpus
0
Project
scv
Extra qsub options
I would like to receive an email when the session starts
Save settings
Save settings and close
Launch
Use these basic options.
Pick one of your projects. Mine is “scv”.
Click “Launch”

Click “Connect to Desktop”



Desktop (399070) 1 core | Running

Host: >_ scc-wl4

Delete

Created at: 2026-02-10 15:56:15 EST

Time Remaining: 11 hours and 59 minutes

Session ID: b6eb15fc-106e-4cf5-ad84-8b2a53125536

List of modules to load (space separated):

Working Directory:

Number of hours: 12

Number of cores: 1

Number of gpus: 0

GPU compute capability: 3.5

Project: scv

Extra qsub options:

Compression

0 (low) to 9 (high)

Image Quality

0 (low) to 9 (high)

Connect to Desktop

View Only (Share-able Link)

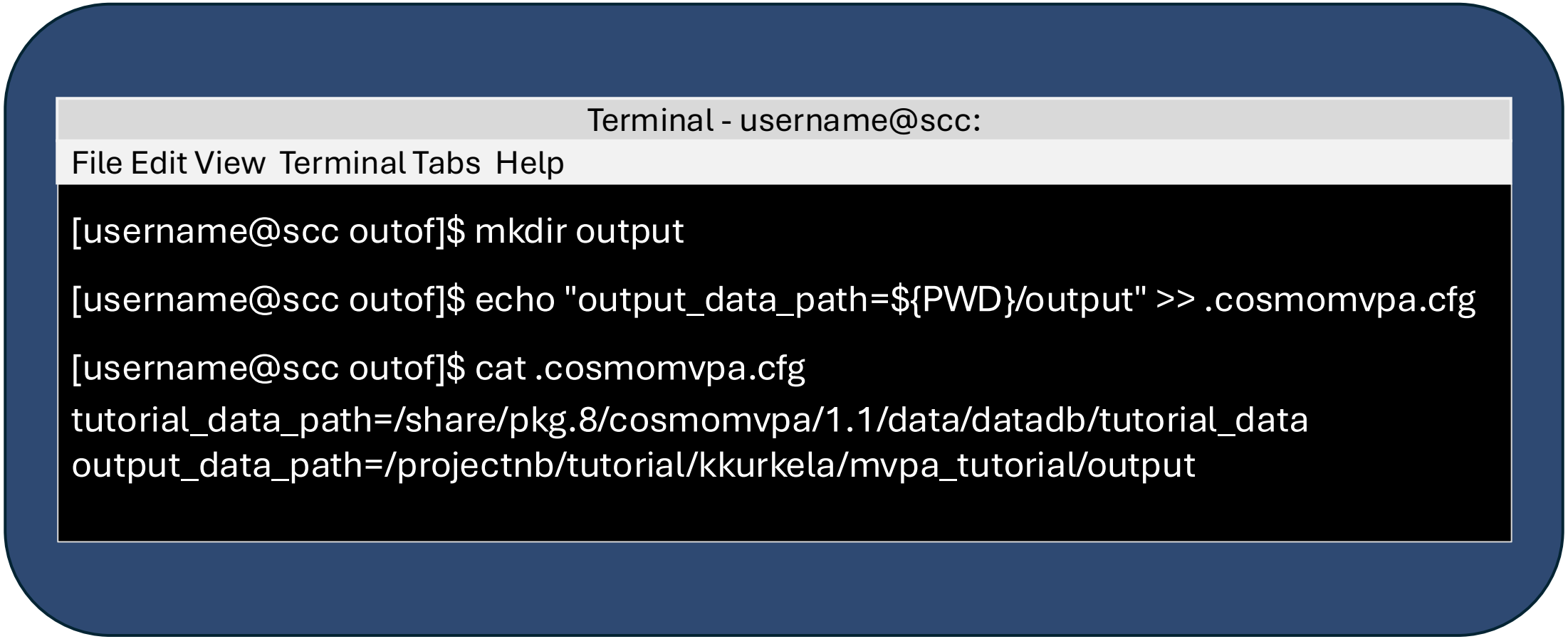
Tutorial Setup

Terminal - username@scc:

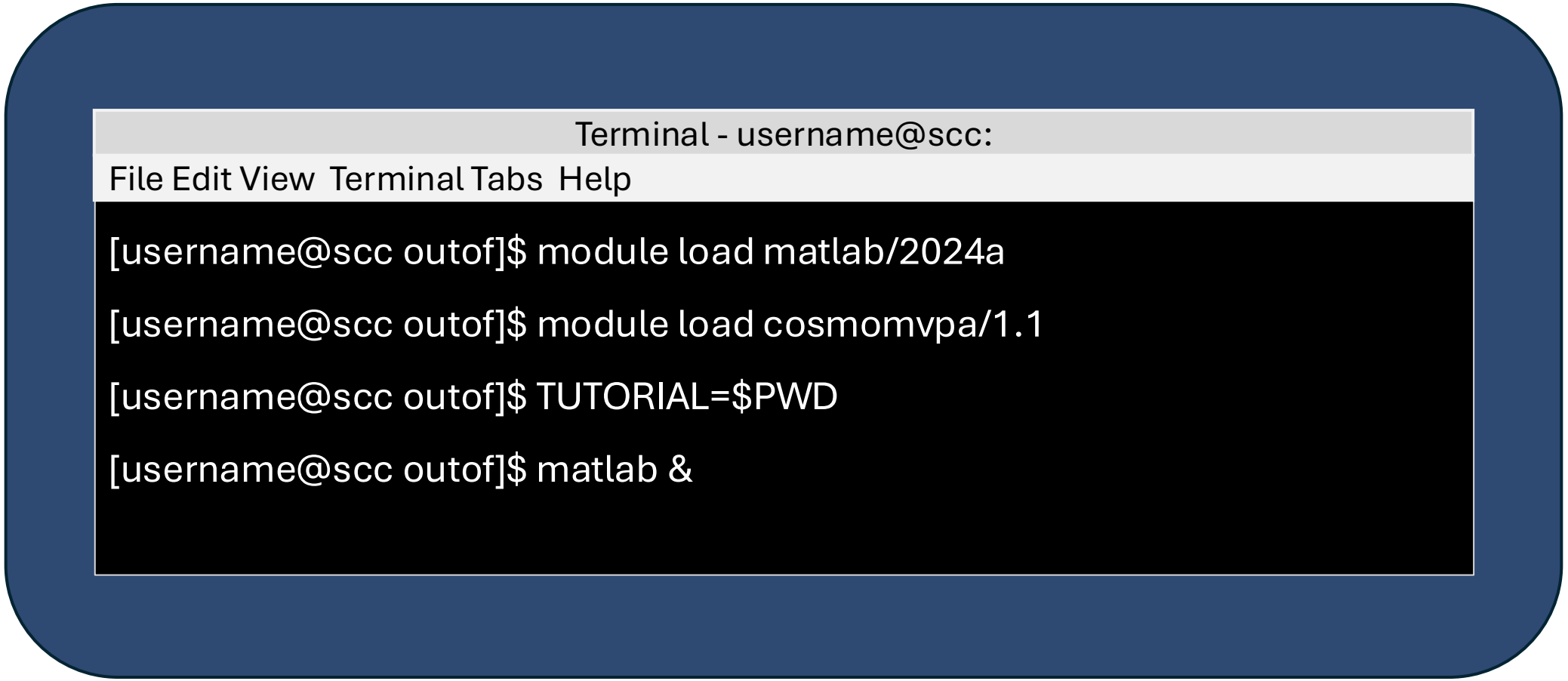
File Edit View Terminal Tabs Help

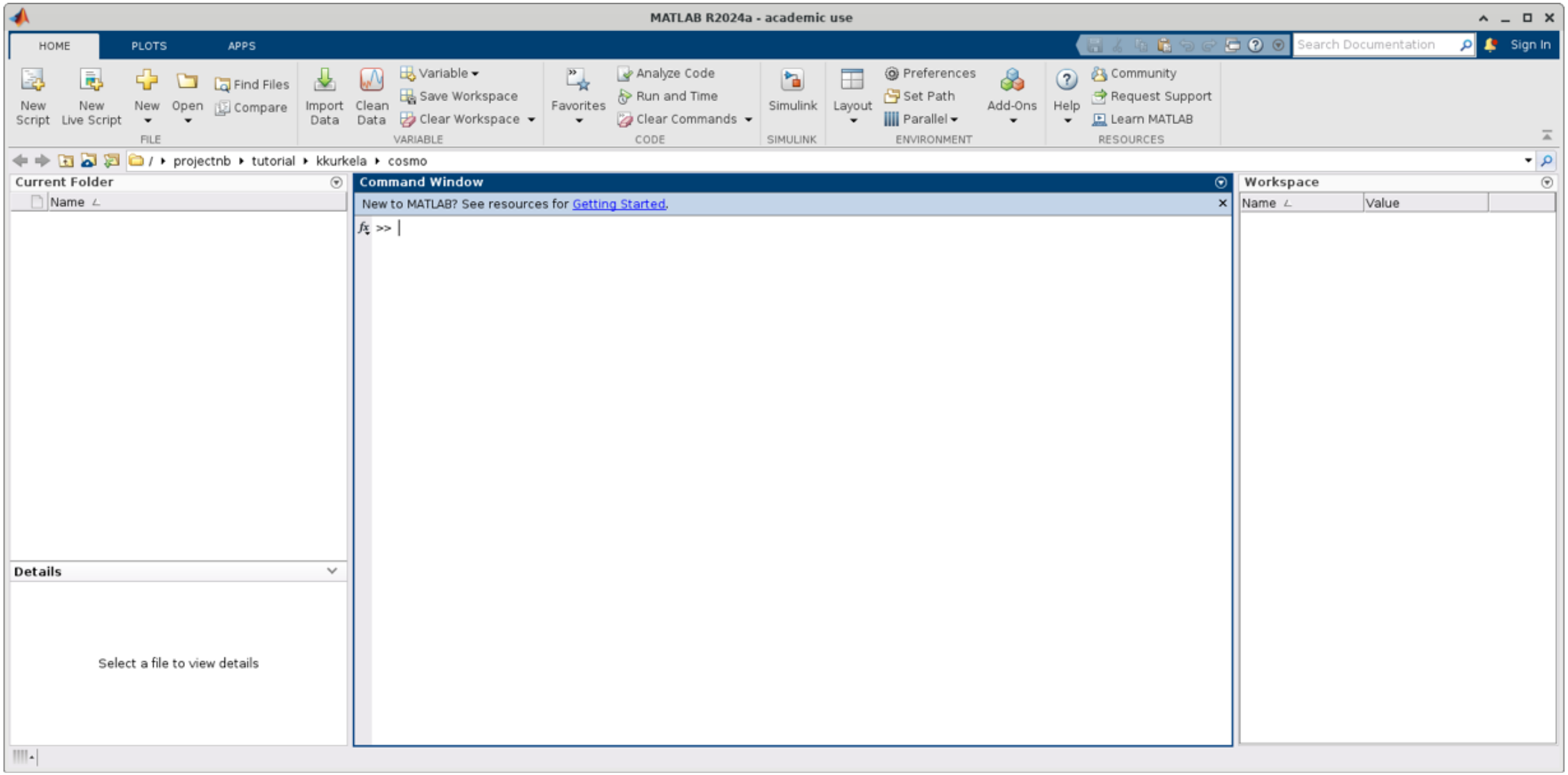
```
[username@scc ~]$ mkdir /path/to/a/directory/to/work/outof
[username@scc outof]$ cd /path/to/a/directory/to/work/outof
[username@scc outof]$ cp –r /project/scv/examples/imaging/tut_mvpa/. .
[username@scc outof]$ ls –a
.   A_load_dataset.m          .cosmomvpa.cfg                D_run_permutation_test.m
..  B_run_nfold_crossvalidate.m  C_run_crossvalidation_searchlight.m  showbrains.sh
```

Tutorial Setup



Tutorial Setup




MATLAB R2024a - academic use
HOME
PLOTS
APPS
Search Documentation
Sign In
New Script
New Live Script
New
Open
Find Files
Compare
FILE
Import Data
Clean Data
Variable
Save Workspace
Clear Workspace
VARIABLE
Favorites
Analyze Code
Run and Time
Clear Commands
CODE
Simulink
SIMULINK
Layout
Preferences
Set Path
Parallel
ENVIRONMENT
Add-Ons
Help
Community
Request Support
Learn MATLAB
RESOURCES
/ ▸ projectnb ▸ tutorial ▸ kkurkela ▸ cosmo
Current Folder
Name
Details
Select a file to view details
Command Window
New to MATLAB? See resources for Getting Started.
>>
Workspace
Name
Value

Outline

- Tutorial Setup
- MVPA Background
- CoSMoMVPA Data Model
- Classification and Cross-Validation
- Searchlights
- Statistical Inference
- Wrap Up

MVPA Background

Univariate Analyses

Imagine the brain was only a 9x9 grid.

We are interested in faces and houses.

Imagine this is the activation that our MRI study revealed.

0	1	0
1	1	1
0	1	0

1	0	1
0	1	0
1	0	1

Univariate Analyses

Typically, we would apply smoothing by "smudging" or "smearing" activation about.

.5	.5	.5
.5	.5	.5
.5	.5	.5

.5	.5	.5
.5	.5	.5
.5	.5	.5

Univariate Analyses

We would then analyze each voxel *independently of the others*.

.5	.5	.5
.5	.5	.5
.5	.5	.5

.5	.5	.5
.5	.5	.5
.5	.5	.5

Univariate Analyses

Is there a difference between faces and house at this voxel?

.5	.5	.5
.5	.5	.5
.5	.5	.5

.5	.5	.5
.5	.5	.5
.5	.5	.5

Univariate Analyses

What about this one?

.5	.5	.5
.5	.5	.5
.5	.5	.5

.5	.5	.5
.5	.5	.5
.5	.5	.5

Univariate Analyses

...and so on

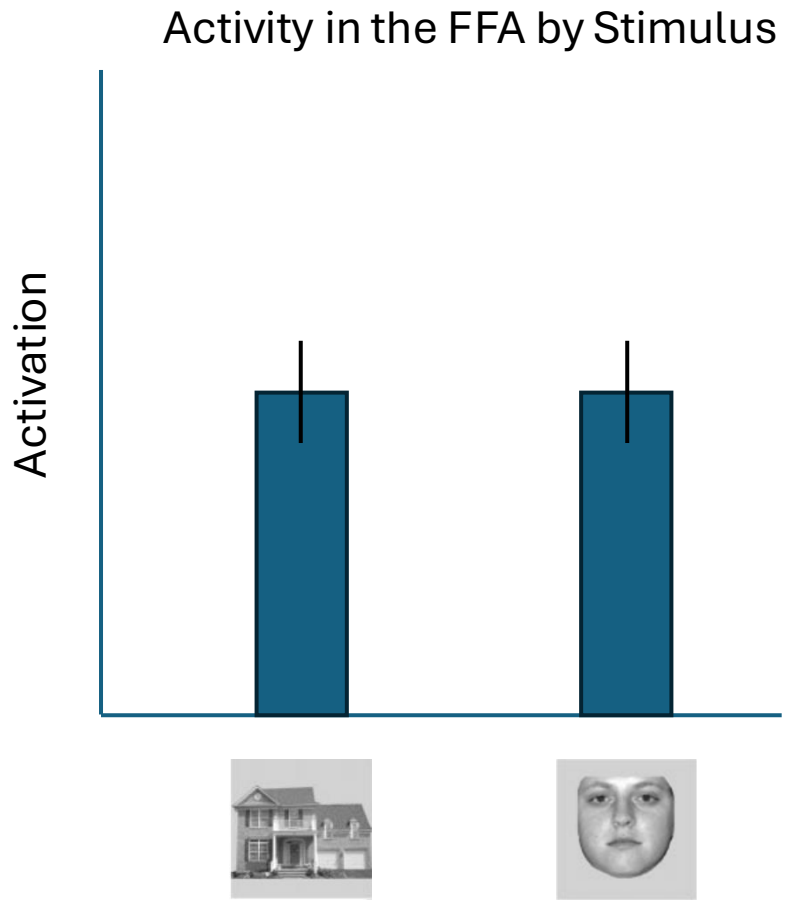
.5	.5	.5
.5	.5	.5
.5	.5	.5

.5	.5	.5
.5	.5	.5
.5	.5	.5

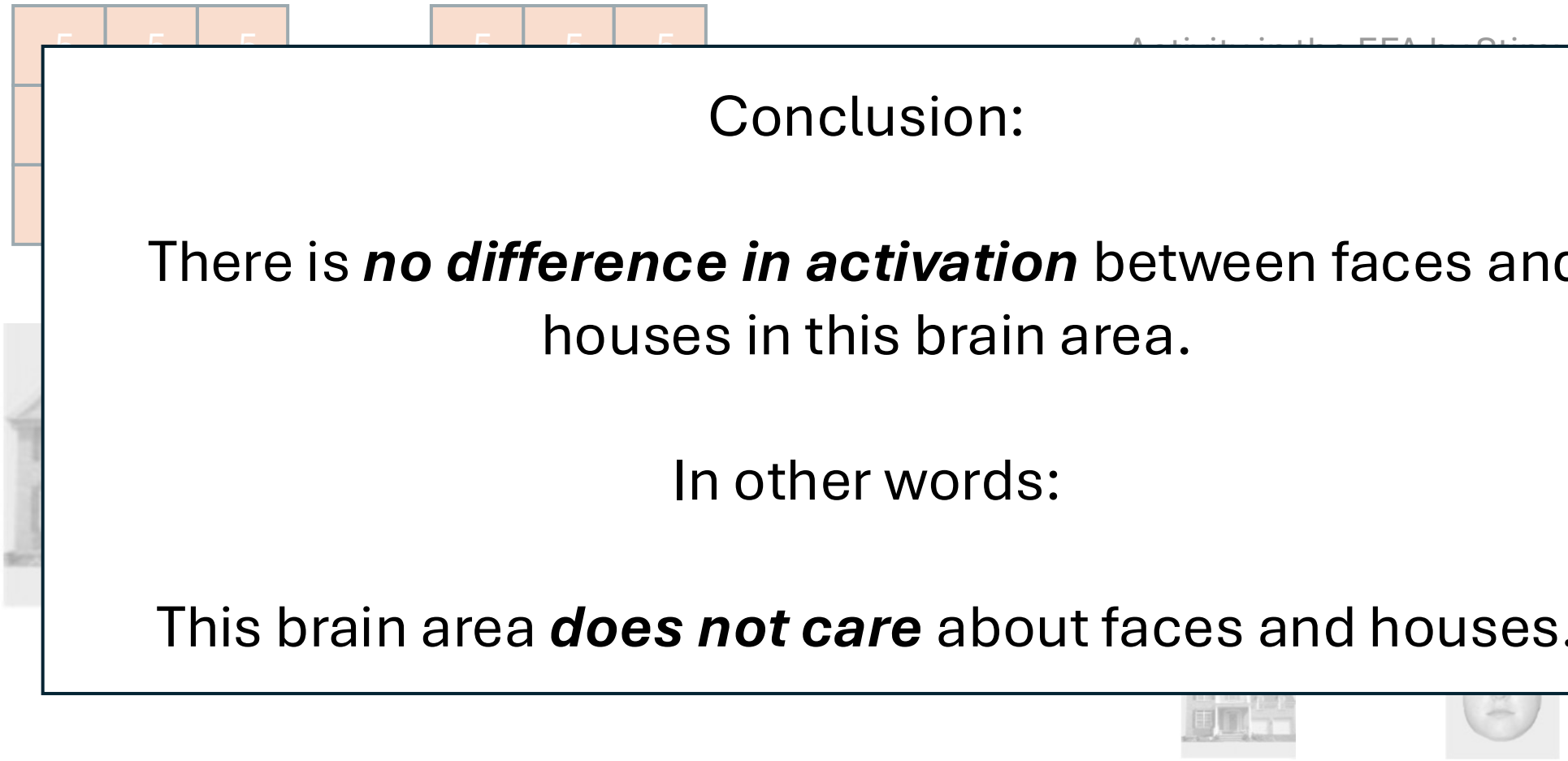
Univariate Analyses

.5	.5	.5
.5	.5	.5
.5	.5	.5

.5	.5	.5
.5	.5	.5
.5	.5	.5



Univariate Analyses



Conclusion:

There is ***no difference in activation*** between faces and houses in this brain area.

In other words:

This brain area ***does not care*** about faces and houses.

~~Univariate~~ Multivariate Analyses

0	1	0
1	1	1
0	1	0

1	0	1
0	1	0
1	0	1

~~Univariate~~ Multivariate Analyses

No Smoothing!

0	1	0
1	1	1
0	1	0

1	0	1
0	1	0
1	0	1

~~Univariate~~ Multivariate Analyses

We consider all the voxels when performing our analysis.

We are looking at the *pattern*.

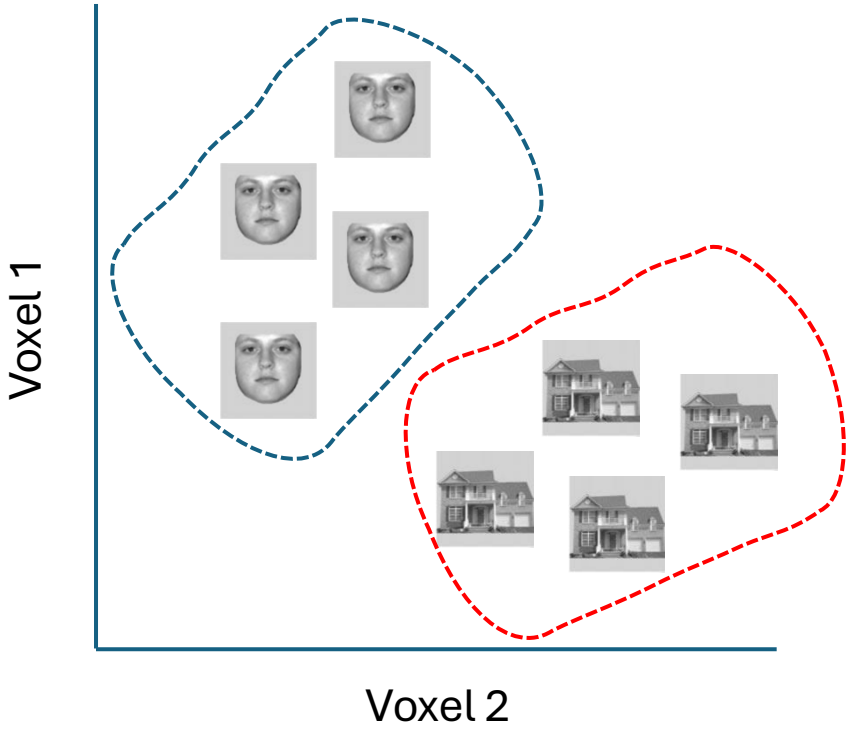
0	1	0
1	1	1
0	1	0

1	0	1
0	1	0
1	0	1

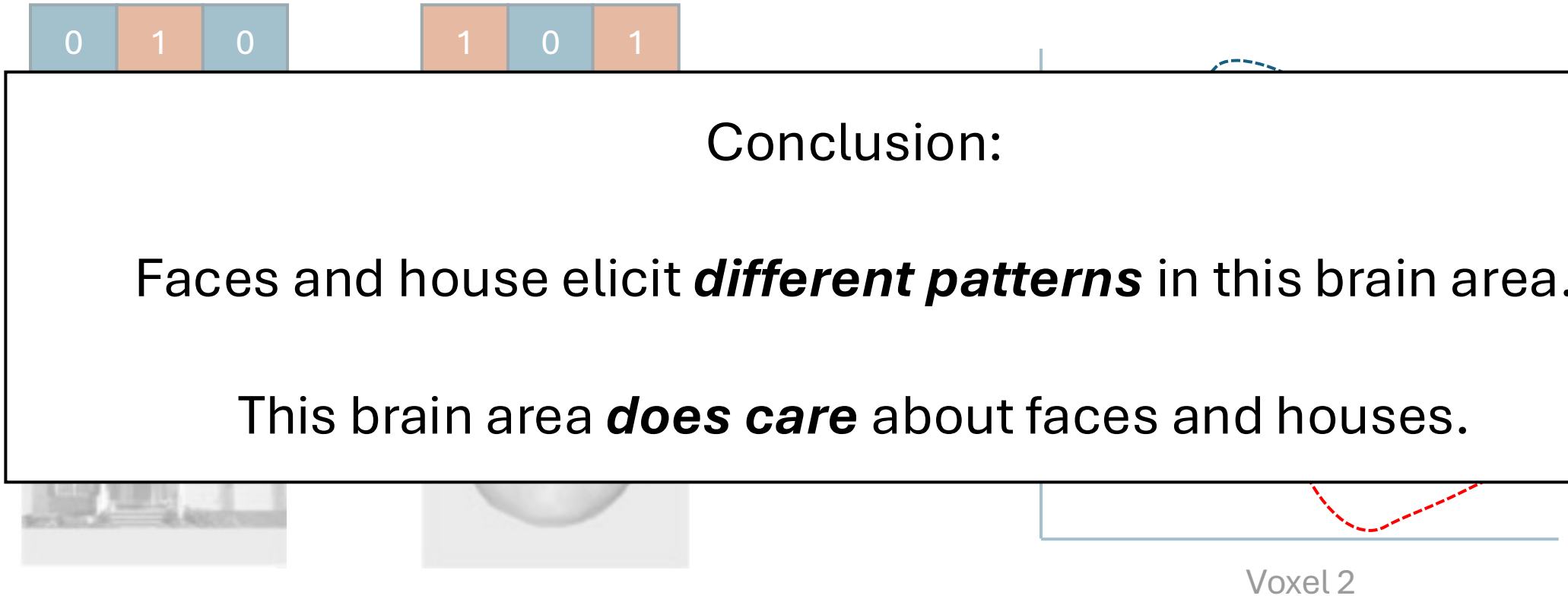
~~Univariate~~ Multivariate Analyses

0	1	0
1	1	1
0	1	0

1	0	1
0	1	0
1	0	1



~~Univariate~~ Multivariate Analyses



Conclusion:

Faces and house elicit ***different patterns*** in this brain area.

This brain area ***does care*** about faces and houses.

Types of Multivariate Analyses

There are different flavors of MVPA. Examples:

- Split-Half Correlation (Haxby et al. 2001 style)
- Decoding/Classification Analyses
- Representational Similarity Analysis (RSA)
- Encoding Models

Data preparation tips for MVPA

- Single Trial / by run statistical estimates
- No (or at least minimal) smoothing!
- Minimizing data distortion
  - Creating subject specific masks
  - Running analyses in subject space
- Maximize the number of samples – the more trials the better

Outline

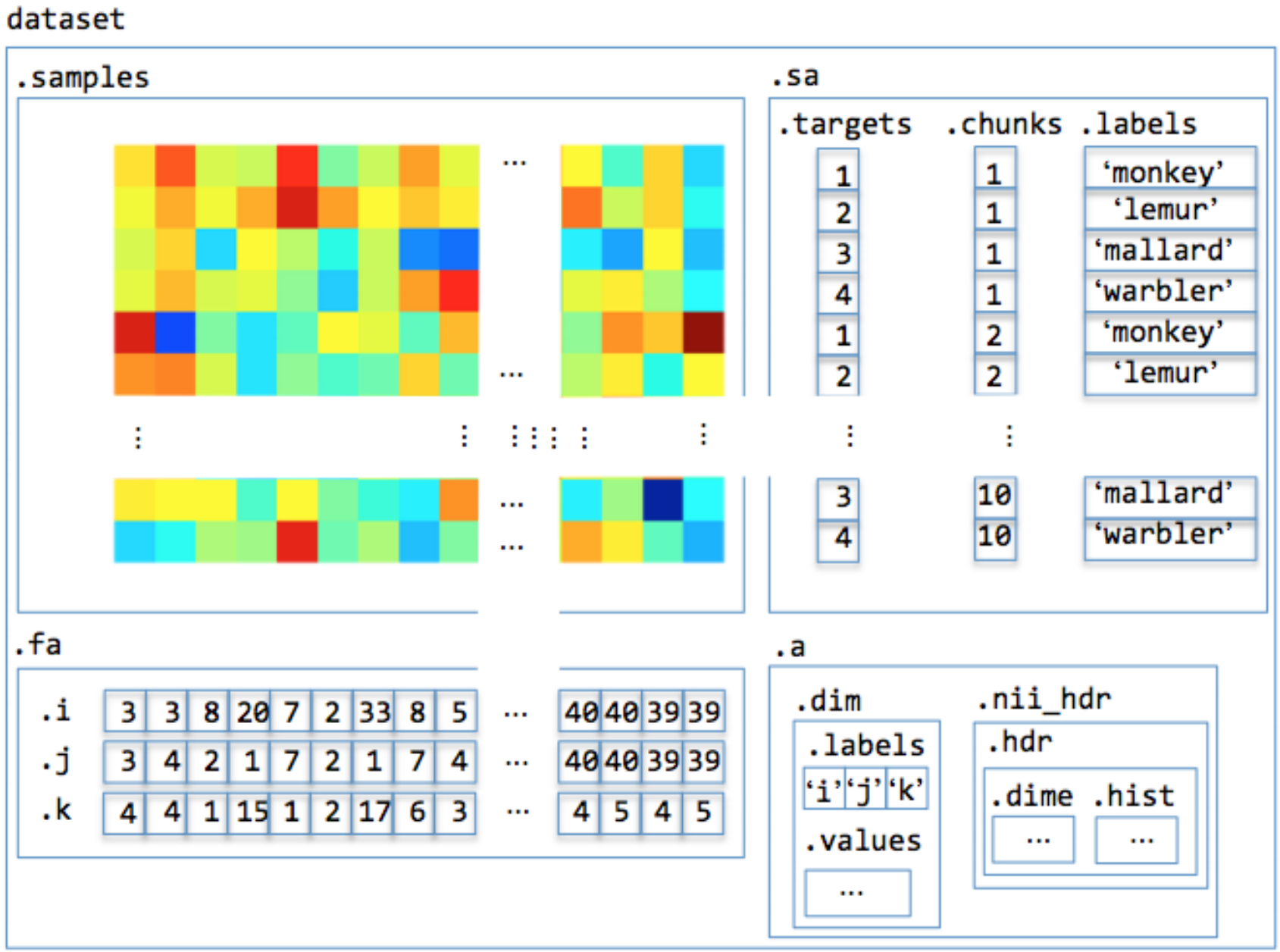
- Tutorial Setup
- MVPA Background
- CoSMoMVPA Data Model
- Classification and Cross-Validation
- Searchlights
- Statistical Inference
- Wrap Up

CoSMo Data Model

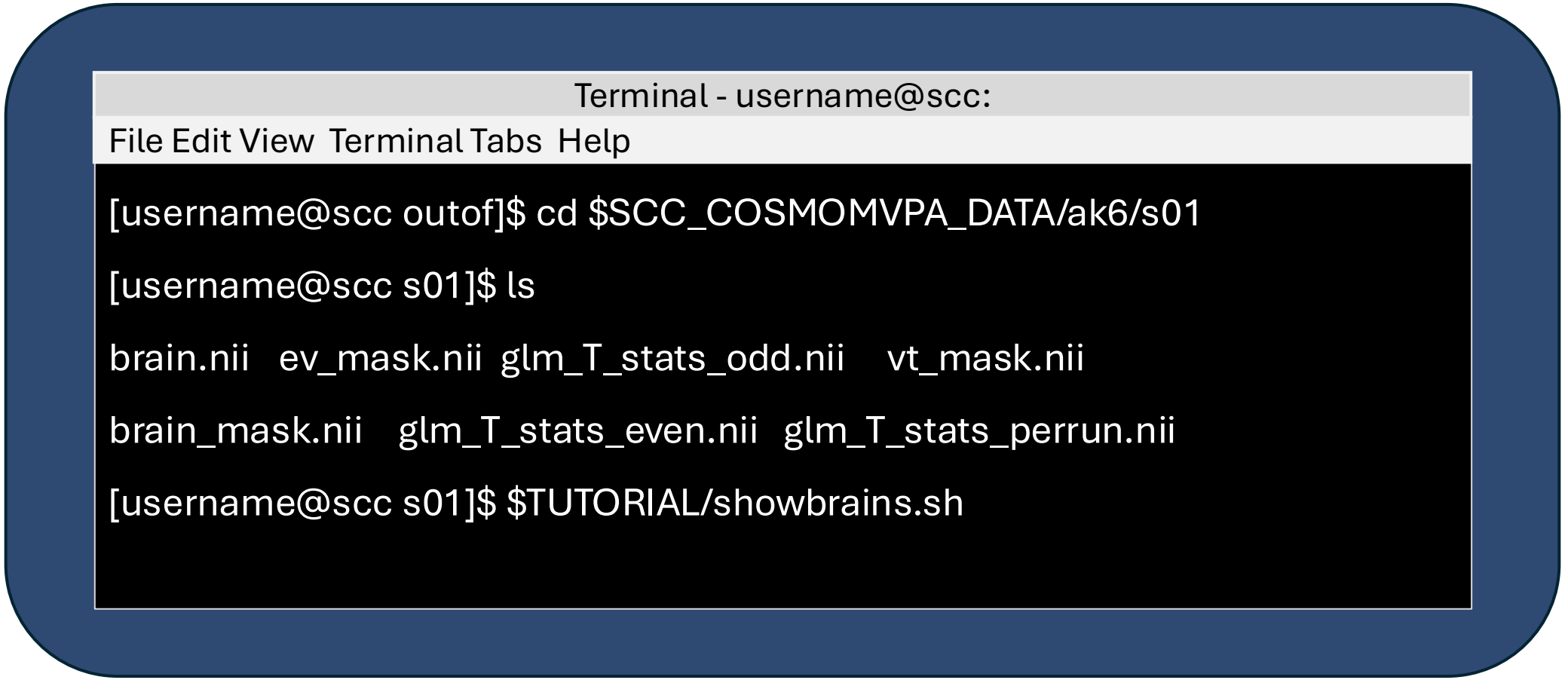
CoSMo Intro

- https://cosmomvpa.org

- CoSMo Concepts
- CoSMo Modules
- CoSMo Demonstrations


dataset
.samples
.sa
.targets
.chunks
.labels
1 1 'monkey'
2 1 'lemur'
3 1 'mallard'
4 1 'warbler'
1 2 'monkey'
2 2 'lemur'
3 10 'mallard'
4 10 'warbler'
.fa
.i 3 3 8 20 7 2 33 8 5 ... 40 40 39 39
.j 3 4 2 1 7 2 1 7 4 ... 40 40 39 39
.k 4 4 1 15 1 2 17 6 3 ... 4 5 4 5
.a
.dim
.labels
'i' 'j' 'k'
.values
...
.nii_hdr
.hdr
.dime
.hist
...
...

Let’s examine the tutorial data




FSLeyes
FSLeyes File Overlay View Settings Tools
brain
3D/4D volume
Opacity
Brightness
Contrast
RESET
Min.
-248.9754
Max.
1529.42
Greyscale
Greyscale
Zoom
100
S
P
A R
S
L R
A
L
P
Overlay list
vt_mask
ev_mask
brain_mask
glm_T_stats_even
brain
Location
Location
Coordinates: Scanner anatomical
Voxel location
-0.677002
125
5.284401
127
54.6237
83
Volume
0
History
Displaying images with different orientations/fields of view!
glm_T_stats_even
[39 39 21 0]: -0.15974807739257812
brain
[125 127 83]: 1289.880615234375

Command Window

```
>> load_dataset
>> cosmo_disp(ds_odd)
.a
  .vol
    .mat
      [ 3       0       0    -122
        0       3       0    -114
        0       0       3   -11.1
        0       0       0       1 ]
    .xform
      'scanner_anat'
    .dim
      [ 80      80      43 ]
  .fdim
    .labels
      { 'i'
        'j'
```

Command Window

f_x

```
% remove all variables defined in the workspace
>> clear

% clear the command window of text
>> clc
```

Outline

- Logistics and Goals
- MVPA Background
- CoSMoMVPA Data Model
- Classification and Cross-Validation
- Searchlights
- Statistical Inference
- Wrap Up

Classification and Cross-validation

What is classification?

Classification is a scientific/statistical/machine learning problem where:

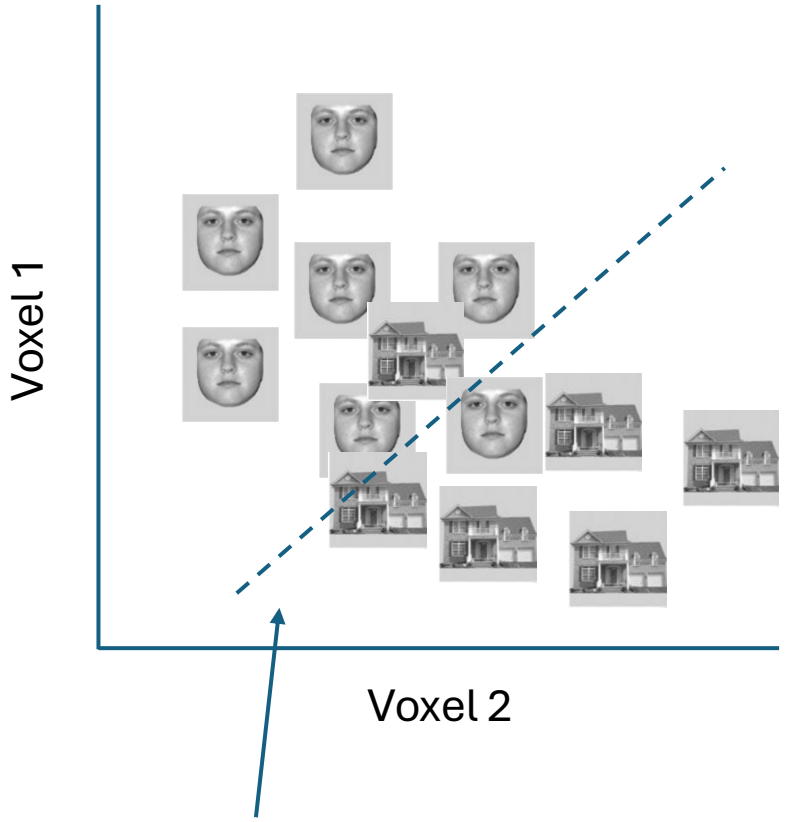
1. Your dataset has 2 or more predictor variables (i.e., voxel 1, voxel 2…voxel n)
2. Your dataset has 1 variable you are interested in predicting (stimulus category)
3. The variable you are interested in predicting is *categorical* (faces vs houses)
4. The variable you are interested in predicting is binary (faces vs houses NOT faces vs houses vs animals)*
5. The variable you are interested in predicting is well balanced (number of houses = number of faces)*
6. Your analytic goal is to build a model that makes accurate predictions out-of-sample (i.e., on new data)

What is classification?

You Data Looks like This:

Voxel 1	Voxel 2	Voxel 3	Voxel n	Stimulus
2.1	-0.5	2.1	-0.5	Face
-0.5	2.1	-0.5	2.1	House
1.3	1.3	1.3	1.3	House
-1.2	1	-1.2	1	Face
1	0.001	1	0.001	Face
0.01	-1.2	0.01	-1.2	House

Can be plotted like this:



Classification is all about drawing the blue dotted line.

There are many types of classifiers

Classifiers are algorithms/models that differ in *how and where* they draw the dotted line.

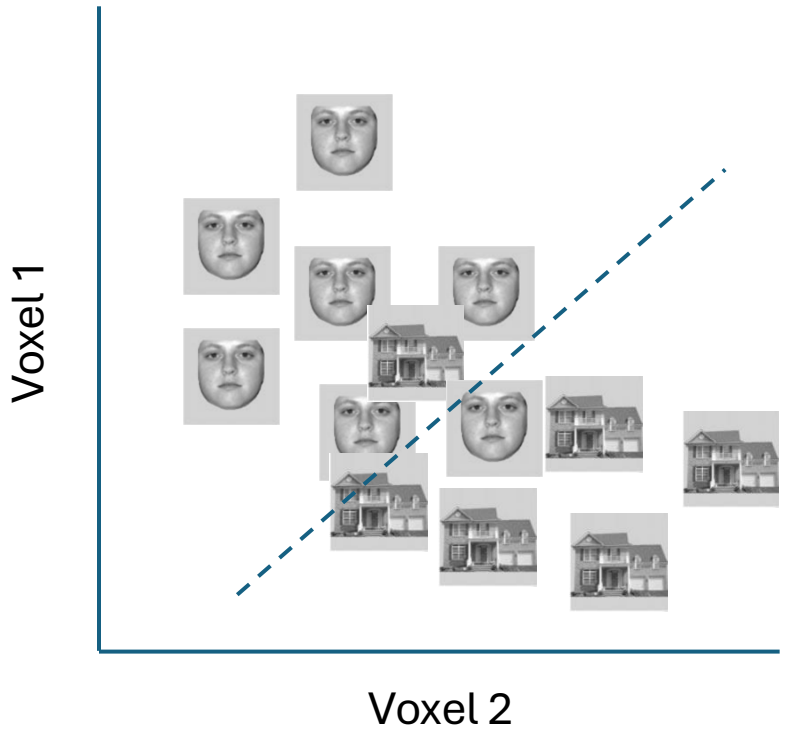
Classifiers come in two types: linear and non-linear.

Each classifier has its strengths and weaknesses.

Classifiers in CoSMoMVPA:

1.) Linear Discriminant Analysis (LDA)
2.) Support Vector Machine (SVM)
3.) Nearest Neighbor (nn)

The classifier we will be using today is the linear classifier LDA.

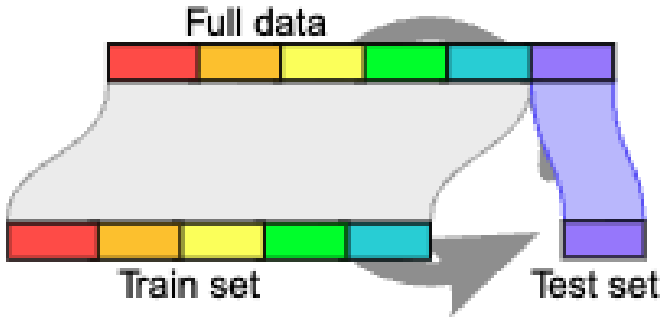


What is cross-validation?

Cross-validation is an algorithm whereby you compute the predictive power of a dataset/classifier combination by splitting your data into *independent chunks* and iteratively folding the data into training sets and test sets.

There are different types of cross validation:

1. N-fold cross validation
2. Leave-one-out cross validation
3. Nested cross validation



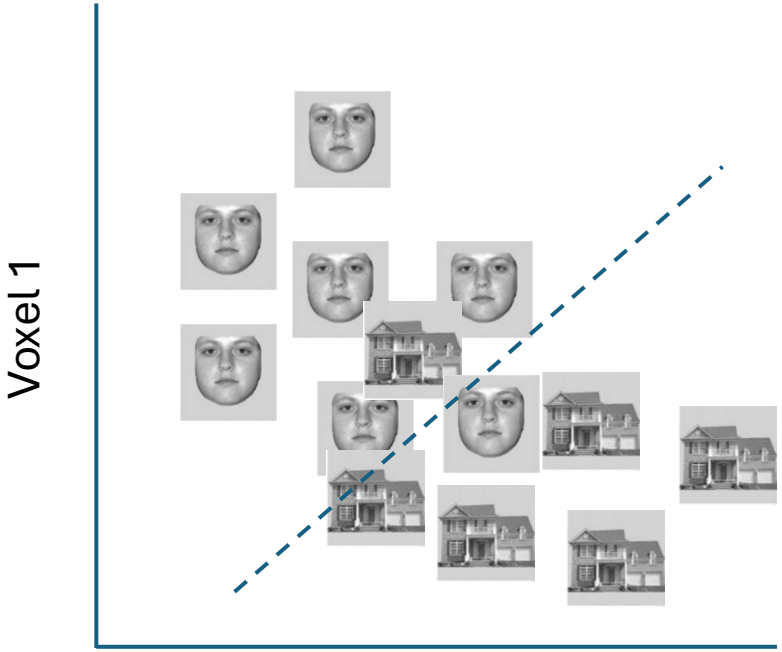
N-fold cross validation

Full Data

Train Model

Test Model

Fold 1



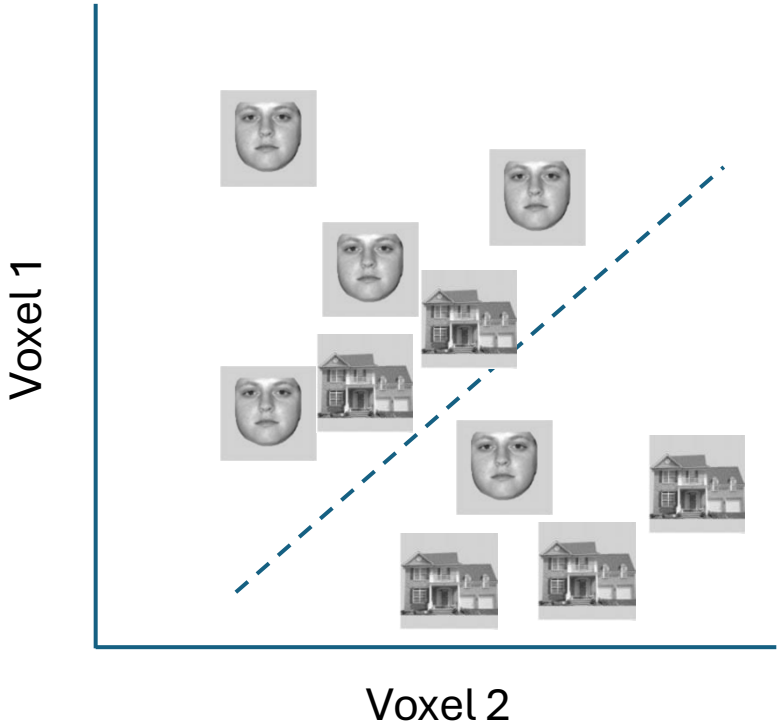
N-fold cross validation

Full Data

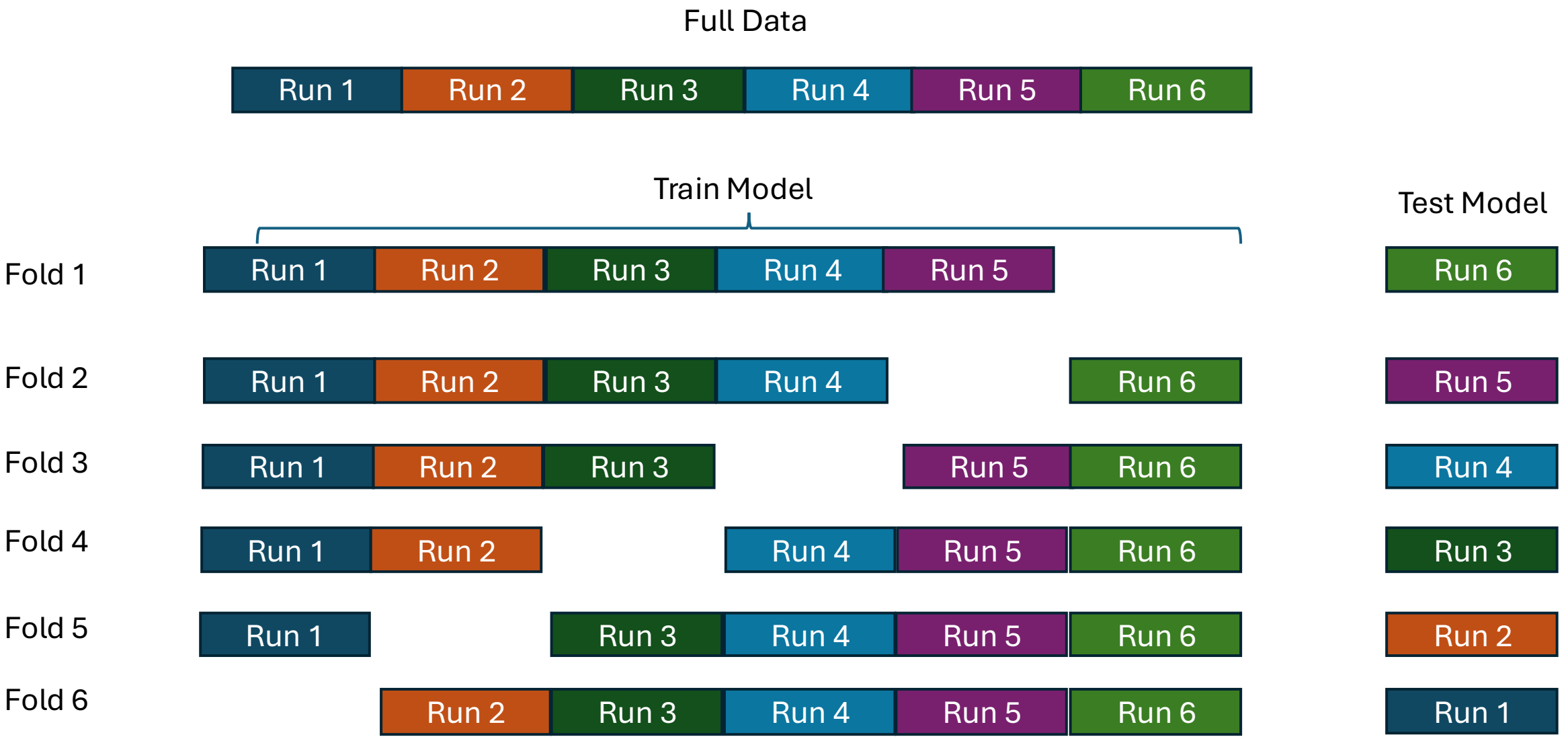
Train Model

Test Model

Fold 1



N-fold cross validation



Command Window

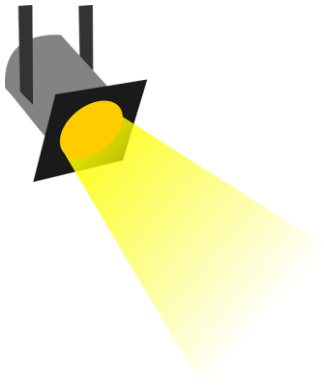
```
% remove all variables defined in the workspace
>> clear

% clear the command window of text
>> clc

% open up the demo script run_nfold_crossvalidate.m
>>  edit run_nfold_crossvalidate.m

% let's run it line-by-line/section by section to understand how the analysis
works
```

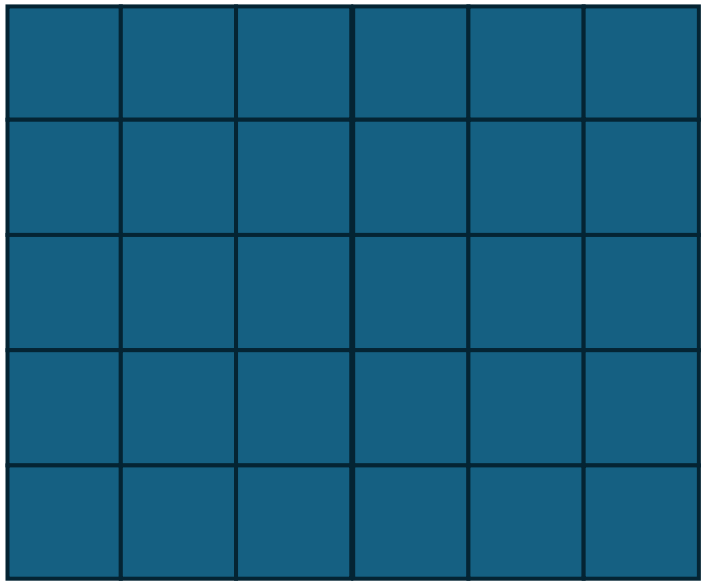
Outline

- Logistics and Goals
- MVPA Background
- CoSMoMVPA Data Model
- Data Prep for MVPA
- Classification and Cross-Validation
- Searchlights
- Statistical Inference
- Wrap Up

Searchlights

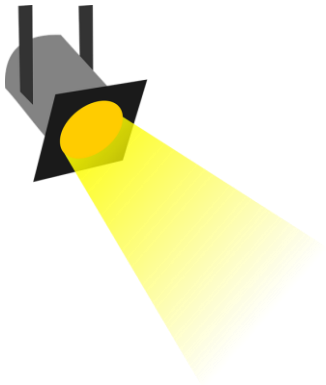
What are searchlights?

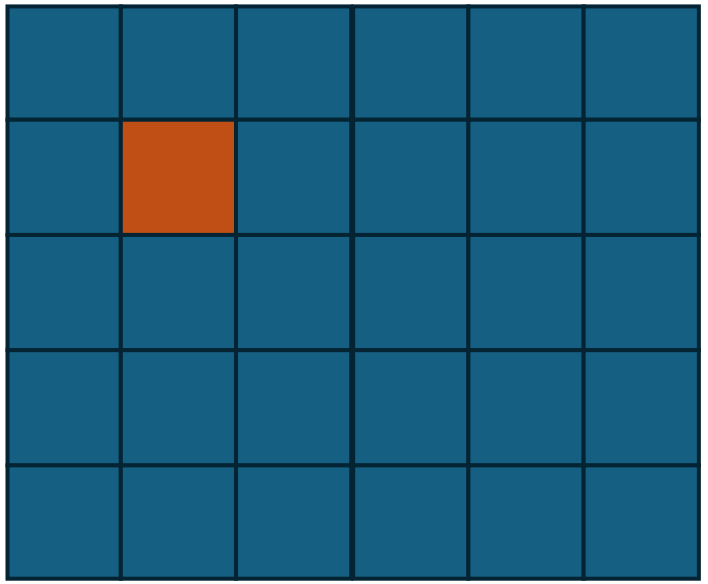




Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868.

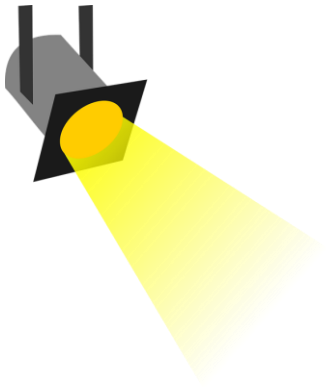
What are searchlights?

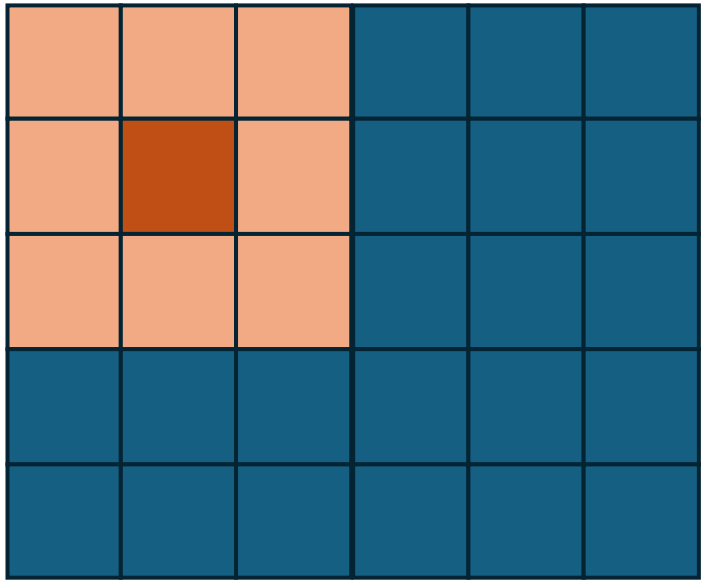




Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868.

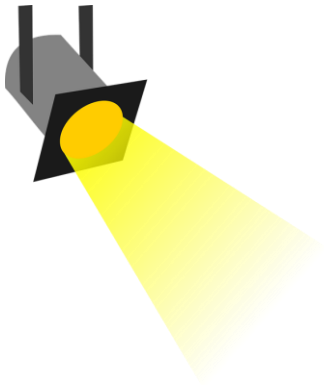
What are searchlights?

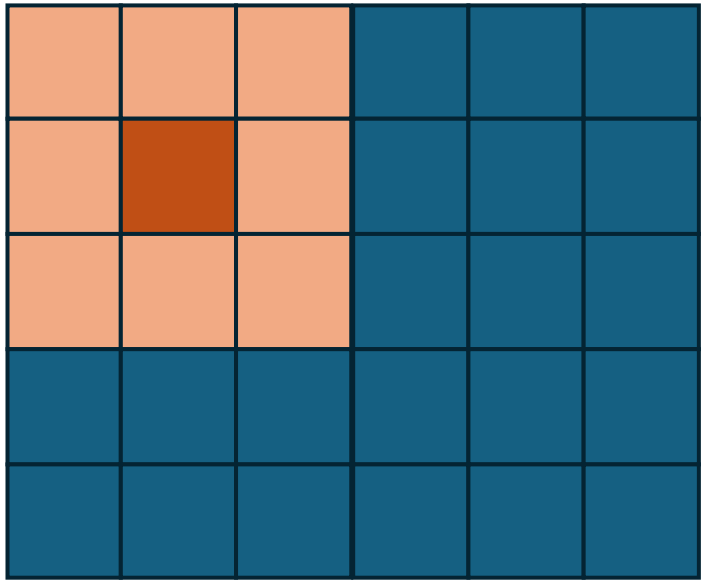


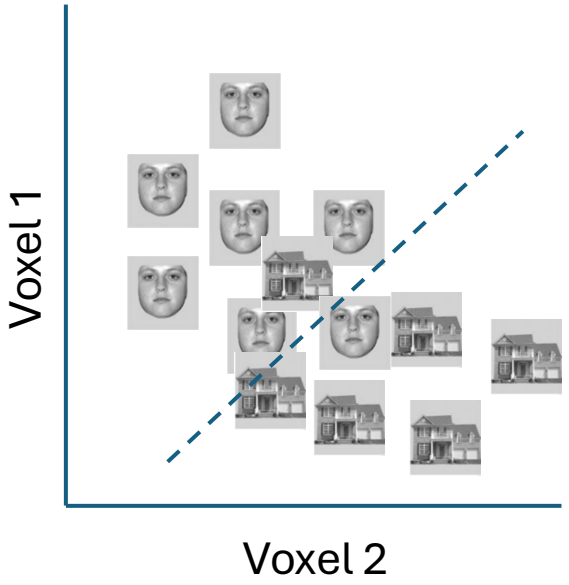


Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868.

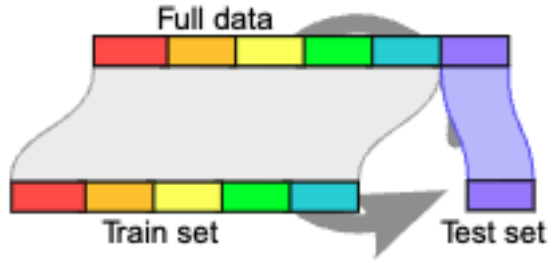
What are searchlights?







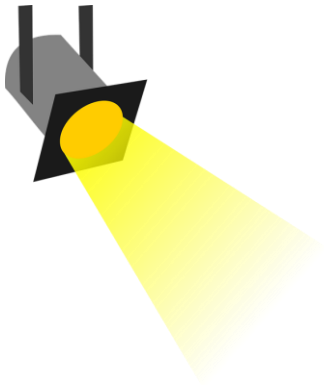
Classification

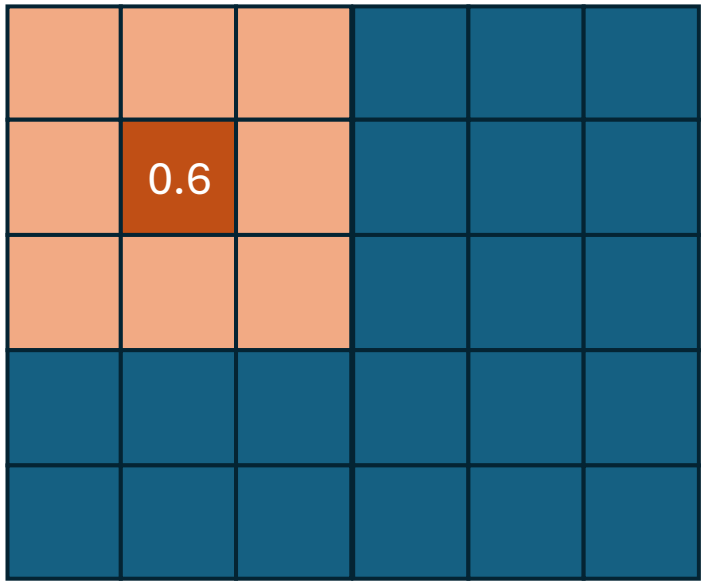


Cross-validation

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868.

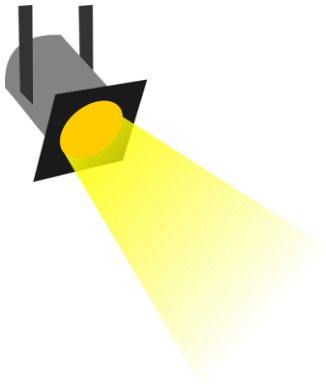
What are searchlights?

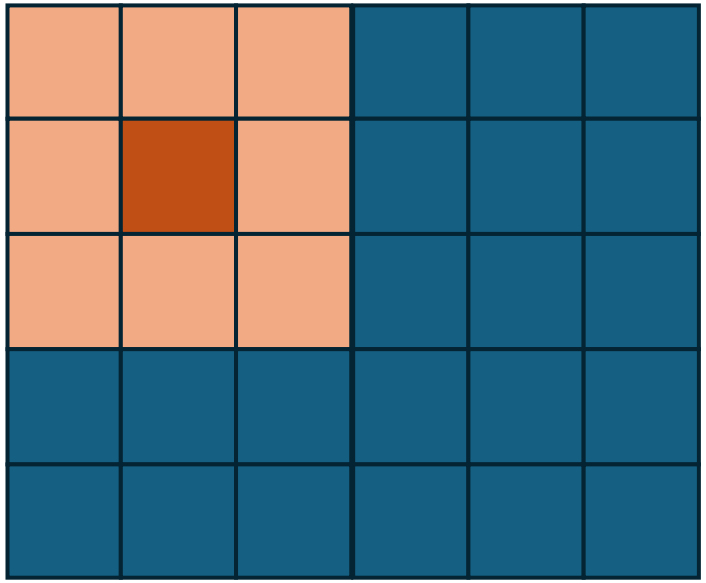




Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868.

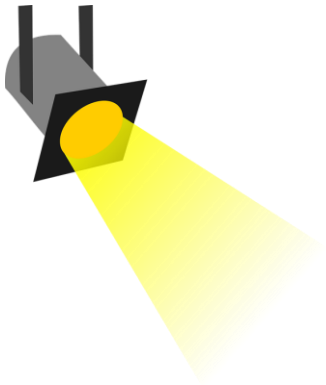
What are searchlights?

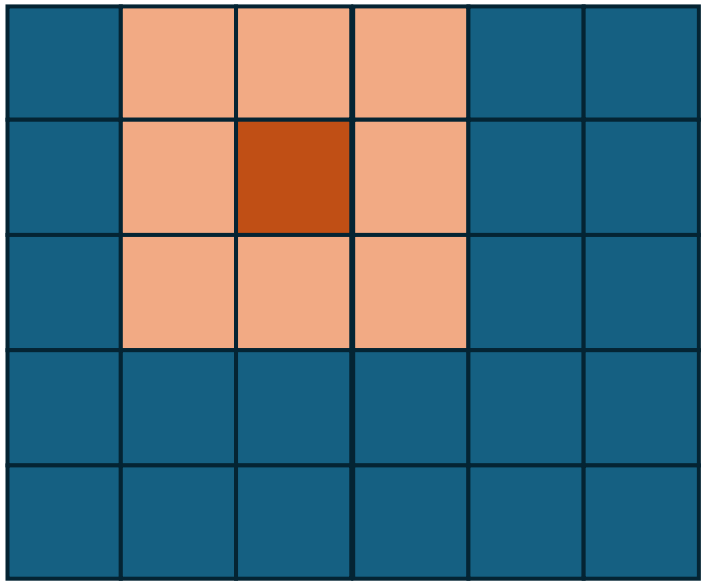




Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868.

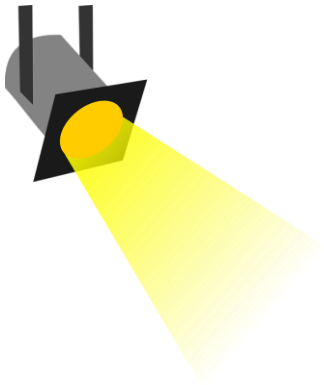
What are searchlights?

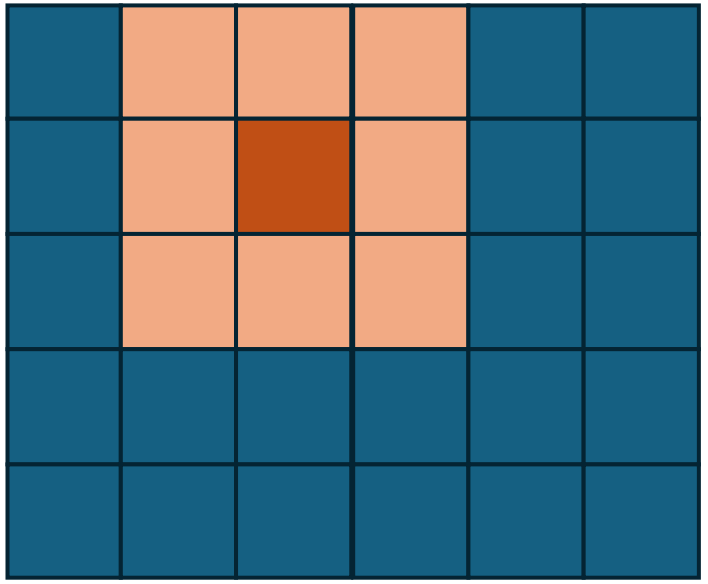


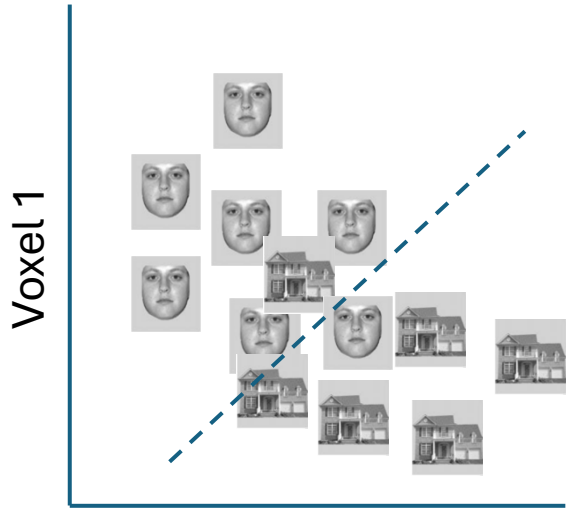


Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868.

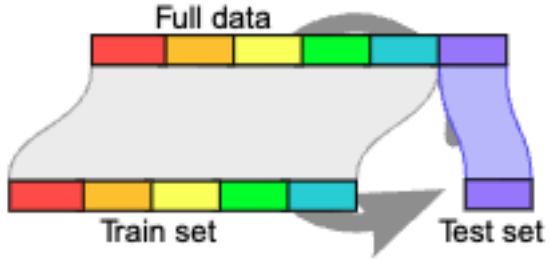
What are searchlights?







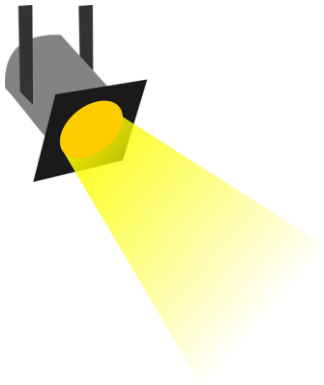
Classification

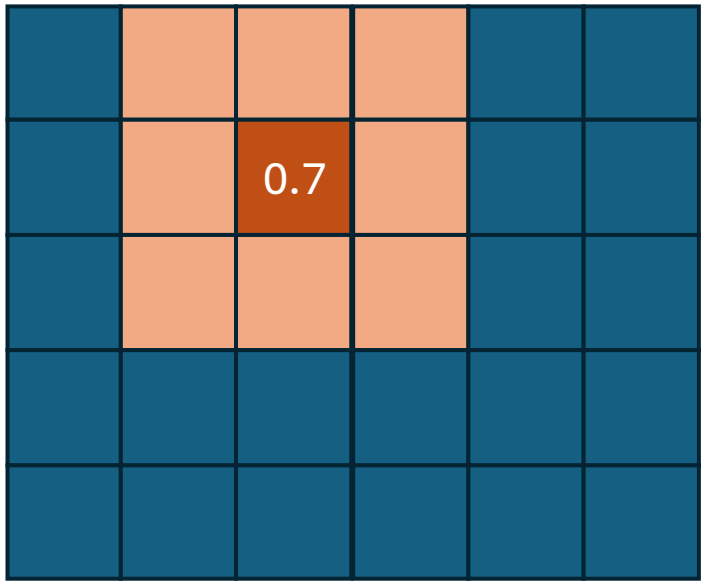


Cross-validation

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868.

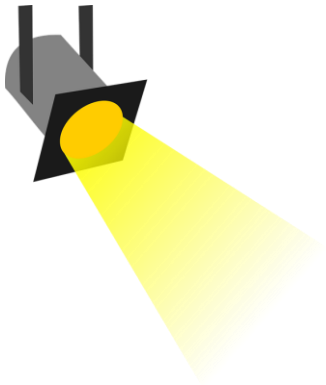
What are searchlights?

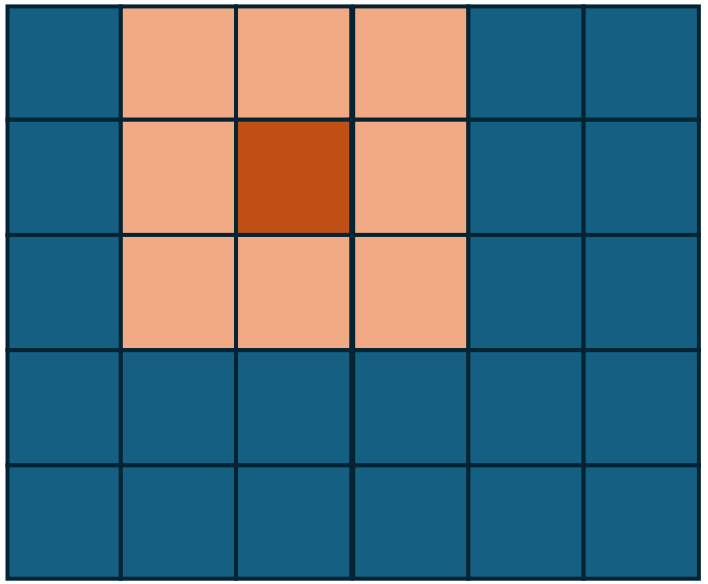




Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868.

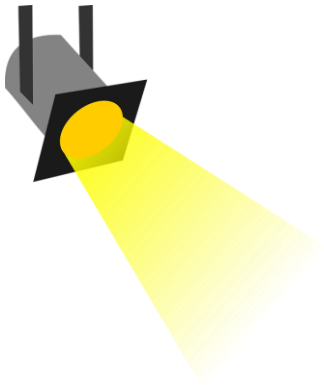
What are searchlights?

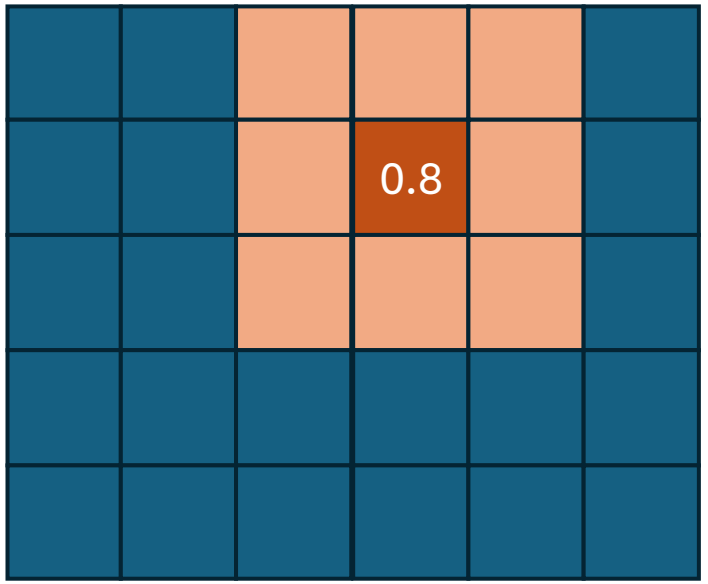




Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868.

What are searchlights?

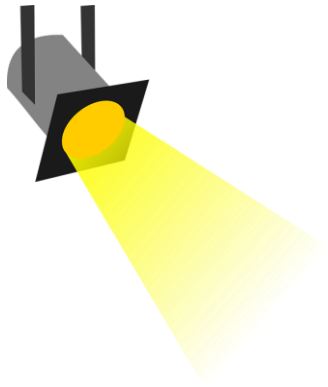


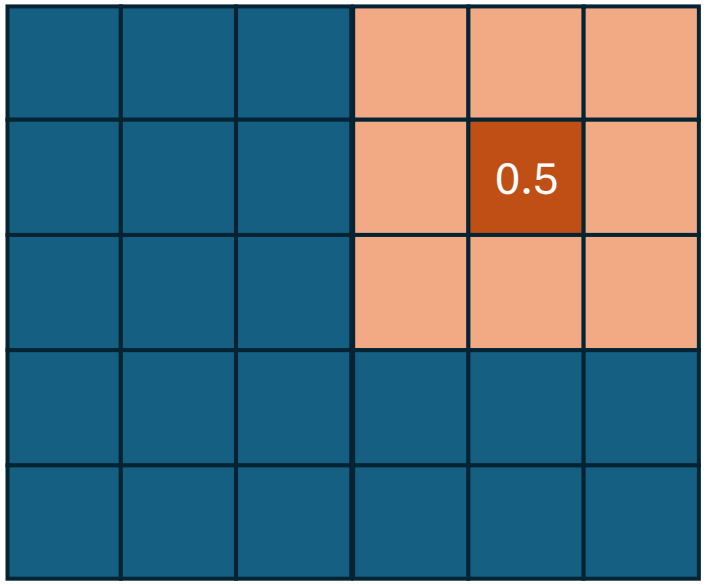


We do the same thing here…

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868.

What are searchlights?

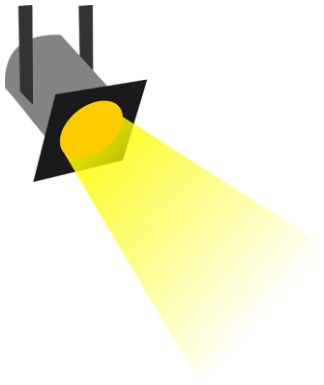


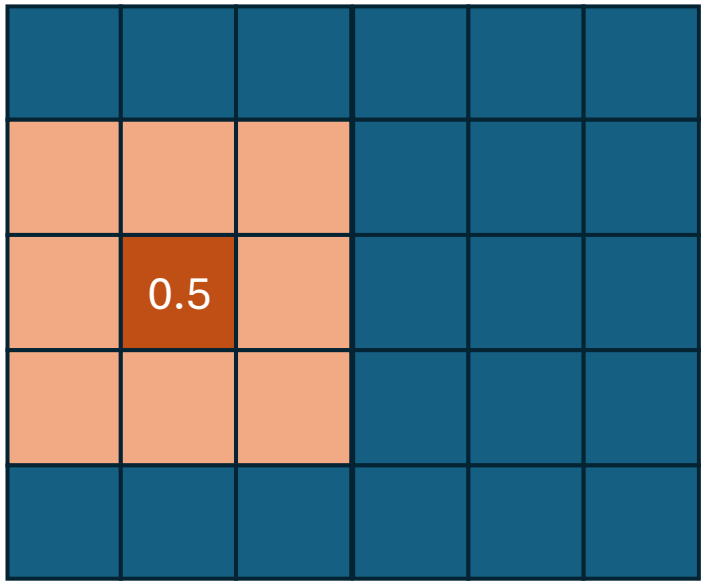


...and here...

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868.

What are searchlights?

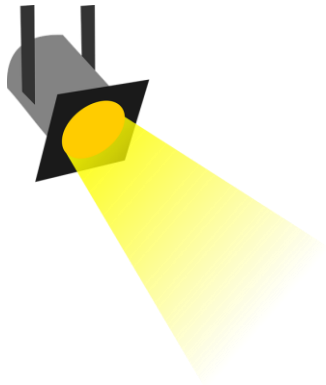


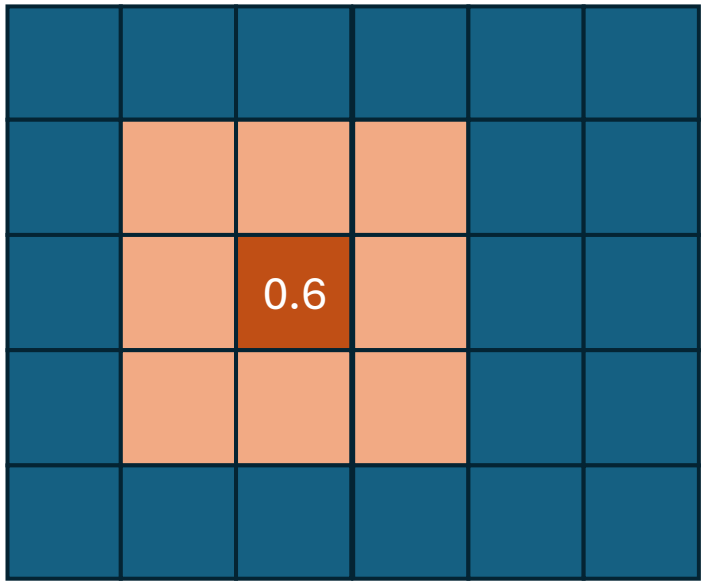


...and here...

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868.

What are searchlights?





...and on in on very every voxel in the brain.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868.

What are searchlights?

- A searchlight is a method where you perform a multivariate analysis (e.g., classification + cross-validation) at every voxel in the brain.
- The method takes each voxel and grabs all voxels within a spherical radius around that voxel, performs the analysis, and then labels the center voxel with the result.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868.

CoSMoMVPA Terminology

- CoSMoMVPA replaces the term “searchlight” with the more general concept of a “neighborhood”.
- Essentially, a spotlight analysis in CoSMoMVPA is like running an ROI analysis 10000s of times – one for each voxel.

Command Window

```
% remove all variables defined in the workspace
>> clear

% clear the command window of text
>> clc

% open up the demo script run_crossvalidation_seachlight.m
>>  edit run_crossvalidation_searchlight.m

% let's run it line-by-line/section by section to understand how the analysis
works
```

Outline

- Logistics and Goals
- MVPA Background
- Intro to the CoSMoMVPA Data Model
- Data Prep for MVPA
- Classification and Cross-Validation
- Searchlights
- Statistical Inference
- Wrap Up

Statistical Inference

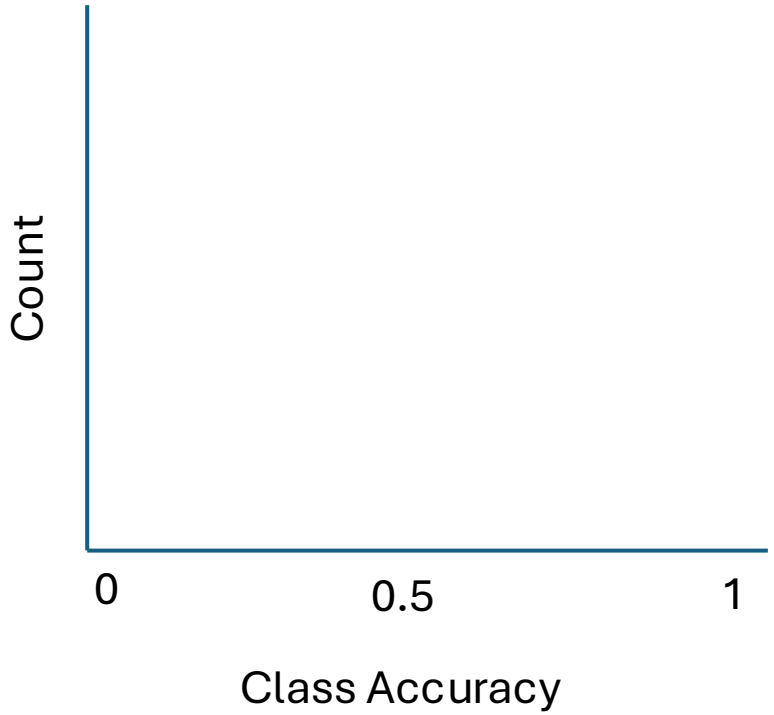
Permutation testing

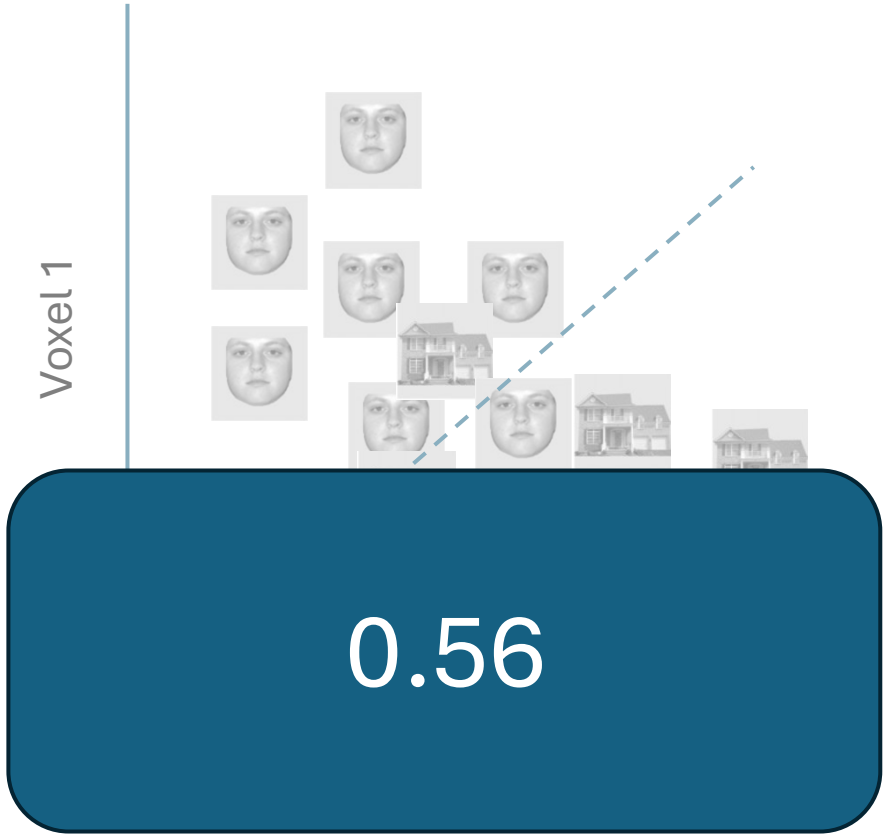
- Practically speaking, real world data are noisy and MRI data classification will result in classification accuracies close to chance.
- Say, for example, brain patterns in the FFA allowed you to classify faces and houses with **51% accuracy**.
- That is above chance – but is it *significantly* so? In other words, could we have gotten this classification accuracy purely due to chance alone?
- How do you determine if you were able to classify “better than chance”, or “significantly”?
- Answer: permutation testing!

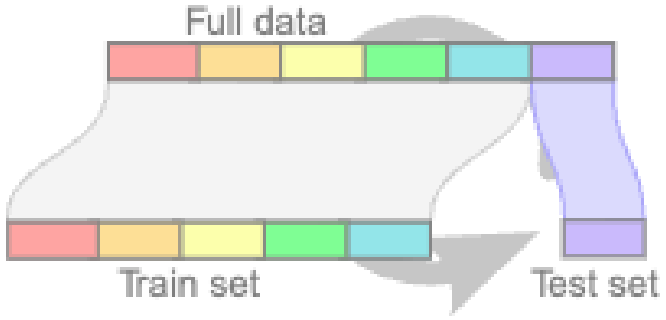
Permutation testing

- *Permutation tests* are a type of non-parametric statistical test for assessing statistical significance.
- Permutation tests try to construct what the distribution of your test statistic (in our case, classification accuracy) looks like under the null hypothesis by **random shuffling trial labels** and rerunning your classification analysis.

Voxel 1	Voxel 2	Voxel 3	Voxel n	Stimulus
2.1	-0.5	2.1	-0.5	Face
-0.5	2.1	-0.5	2.1	House
1.3	1.3	1.3	1.3	House
-1.2	1	-1.2	1	Face
1	0.001	1	0.001	Face
0.01	-1.2	0.01	-1.2	House

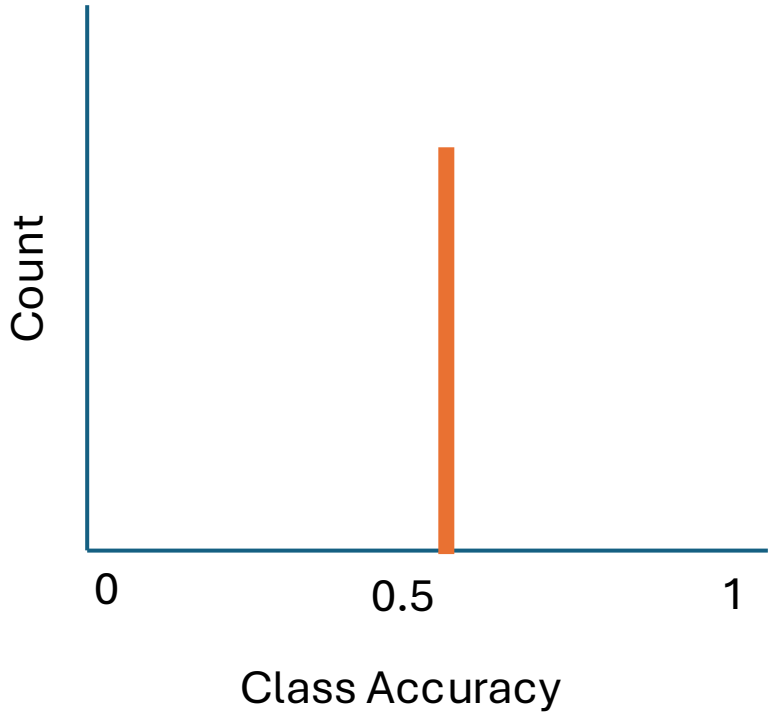




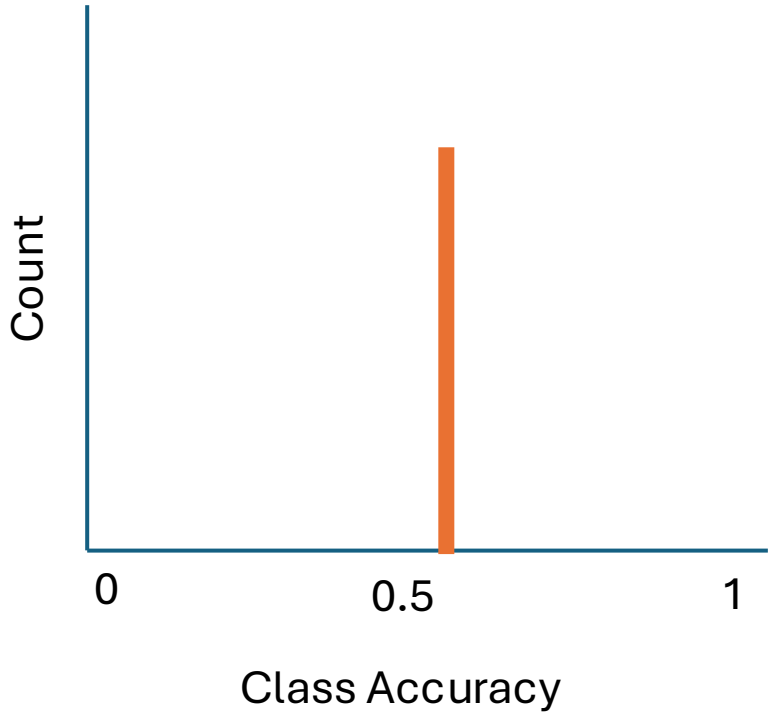


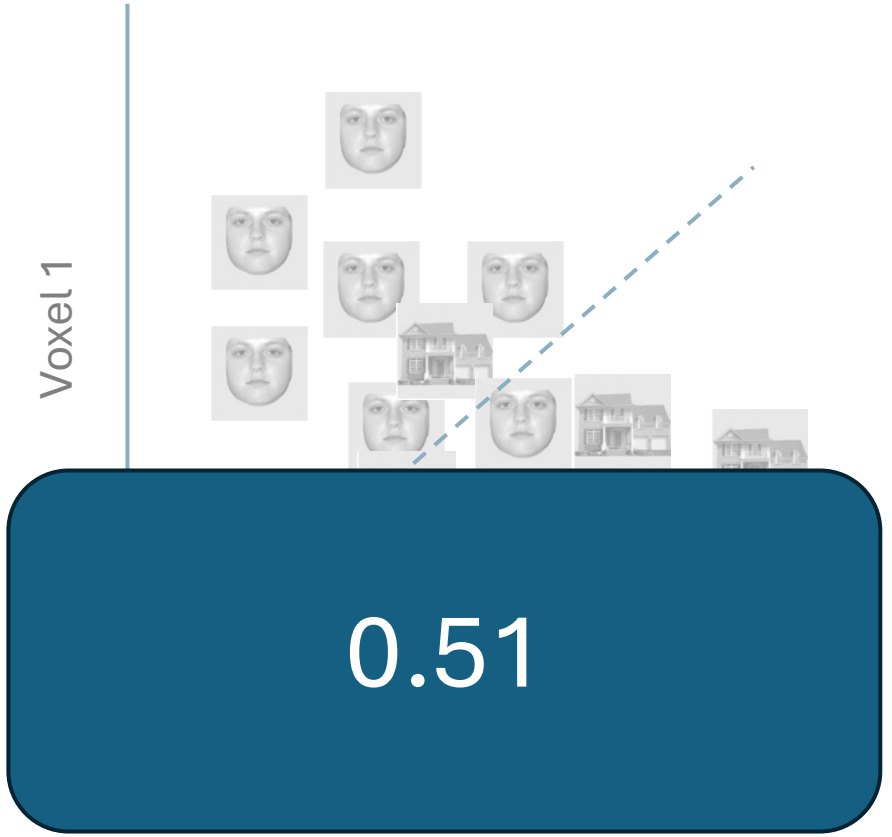
Cross-validation

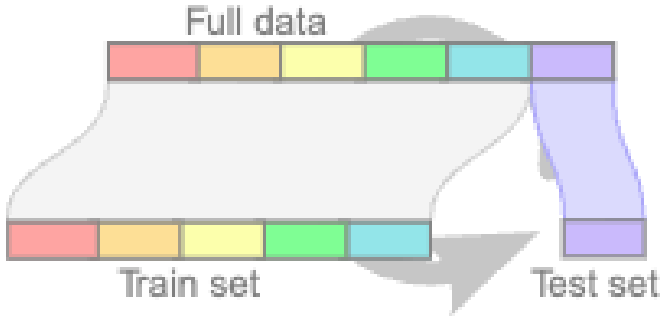
Voxel 1	Voxel 2	Voxel 3	Voxel n	Stimulus
2.1	-0.5	2.1	-0.5	Face
-0.5	2.1	-0.5	2.1	House
1.3	1.3	1.3	1.3	House
-1.2	1	-1.2	1	Face
1	0.001	1	0.001	Face
0.01	-1.2	0.01	-1.2	House



Voxel 1	Voxel 2	Voxel 3	Voxel n	Stimulus
2.1	-0.5	2.1	-0.5	House
-0.5	2.1	-0.5	2.1	House
1.3	1.3	1.3	1.3	Face
-1.2	1	-1.2	1	Face
1	0.001	1	0.001	House
0.01	-1.2	0.01	-1.2	Face



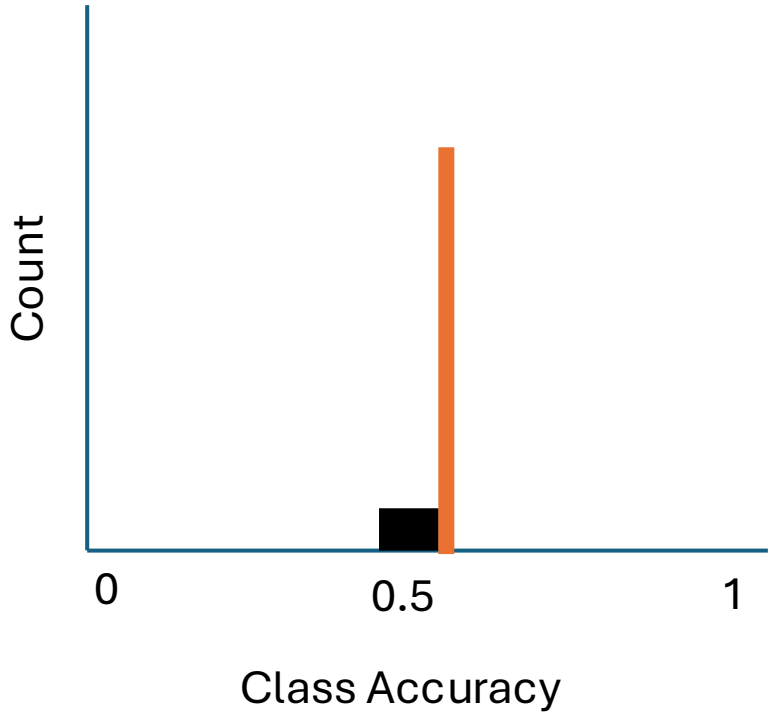




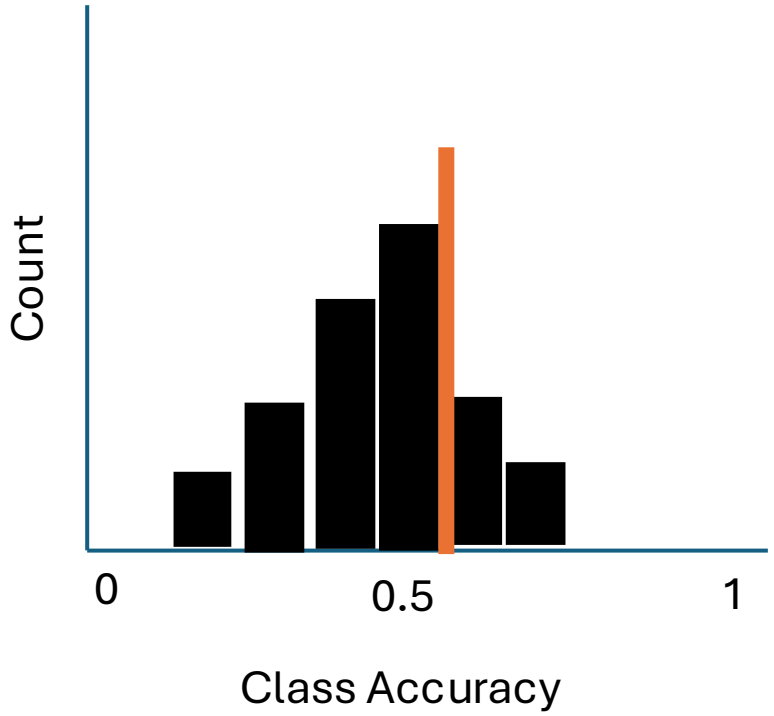
Cross-validation

Voxel 1	Voxel 2	Voxel 3	Voxel n	Stimulus
2.1	-0.5	2.1	-0.5	House
-0.5	2.1	-0.5	2.1	House
1.3	1.3	1.3	1.3	Face
-1.2	1	-1.2	1	Face
1	0.001	1	0.001	House
0.01	-1.2	0.01	-1.2	Face

Repeat 1000x



Voxel 1	Voxel 2	Voxel 3	Voxel n	Stimulus
2.1	-0.5	2.1	-0.5	House
-0.5	2.1	-0.5	2.1	House
1.3	1.3	1.3	1.3	Face
-1.2	1	-1.2	1	Face
1	0.001	1	0.001	House
0.01	-1.2	0.01	-1.2	Face



The orange line (observed accuracy) is greater than X% of null simulations.

The p-value of this test would be 1-X.

Command Window

```
% remove all variables defined in the workspace
>> clear

% clear the command window of text
>> clc

% open up the demo script run_permutation_test.m
>>  edit run_permutation_test.m

% let's run it line-by-line/section by section to understand how the analysis
works
```

Outline

- Logistics and Goals
- MVPA Background
- Intro to the CoSMoMVPA Data Model
- Data Prep for MVPA
- Classification and Cross-Validation
- Searchlights
- Statistical Inference
- Wrap Up

Tutorial Feedback

Please open a tab in your web browser and have this page open:

http://scv.bu.edu/survey/tutorial_evaluation.html

Please fill it out at the end of todays tutorial!

Questions? Comments? Concerns?