OHBM
Brainhack
2020

Introduction to ML / sklearn / nilearn

Michael Dayan

Methods & Data facility
Human Neuroscience Platform
Foundation Campus Biotech Geneva

HNP METHODS & DATA

All content at: https://github.com/mick-d/introml-sklearn-nilearn

campus biotech

Fondation Campus Biotech Geneva

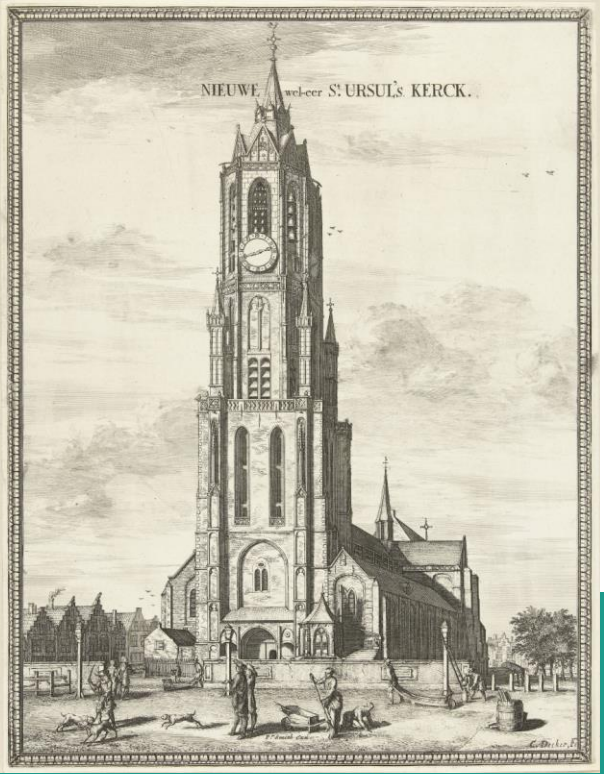

Classification
General examples about classification algorithms.
Normal and Shrinkage Linear Discriminant Analysis for classification
Recognizing hand-written digits
Plot classification probability
Classifier comparison
Linear and Quadratic Discriminant Analysis with covariance ellipsoid
Machine Learning (Decoding, Encoding, and MVPA)
Decoding, encoding, and general machine learning examples.
Representational Similarity Analysis
Motor imagery decoding from EEG data using the Common Spatial Pattern (CSP)
Decoding in time-frequency space data using the Common Spatial Pattern (CSP)
Decoding source space data
MNE
MEG + EEG ANALYSIS & VISUALIZATION
Clustering
Examples concerning the sklearn.cluster module.
scikit learn
8.1. Tutorial examples
Introductory examples that teach how to use nilearn.
Basic numerics and plotting with Python
Basic nilearn example: manipulating and looking at data
3D and 4D niimgs: handling and visualizing
A introduction tutorial to fMRI decoding
8.2. Visualization of brain images
See Plotting brain images for more details.
Basic Atlas plotting
Glass brain plotting in nilearn
Visualizing Megatrawls Network Matrices from Human Connectome Project
Visualizing multiscale functional brain parcellations
Technical point: Illustration of the volume to surface sampling
ni learn

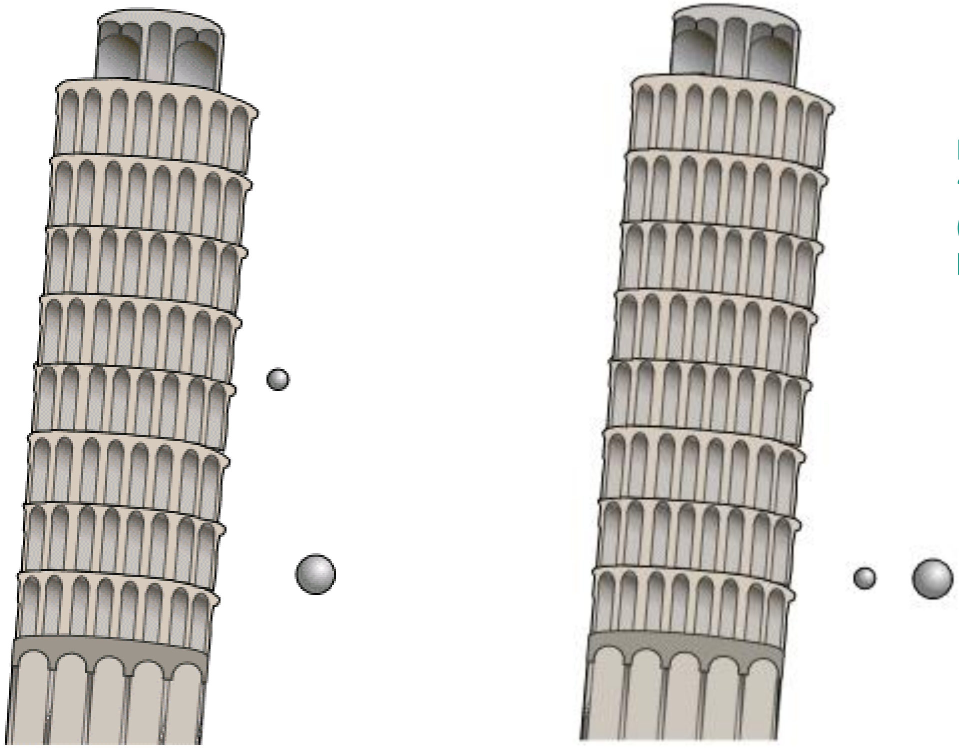
SOME HALLMARKS OF SCIENCE

- Testable predictions

Aristotle's theory of gravity: objects fall at a speed proportional to their mass

Galileo's theory of gravity: objects fall at the same speed





Aristotle's theory

Galileo's theory

Figure adapted from "Layers of Learning" (https://layers-of-learning.com)

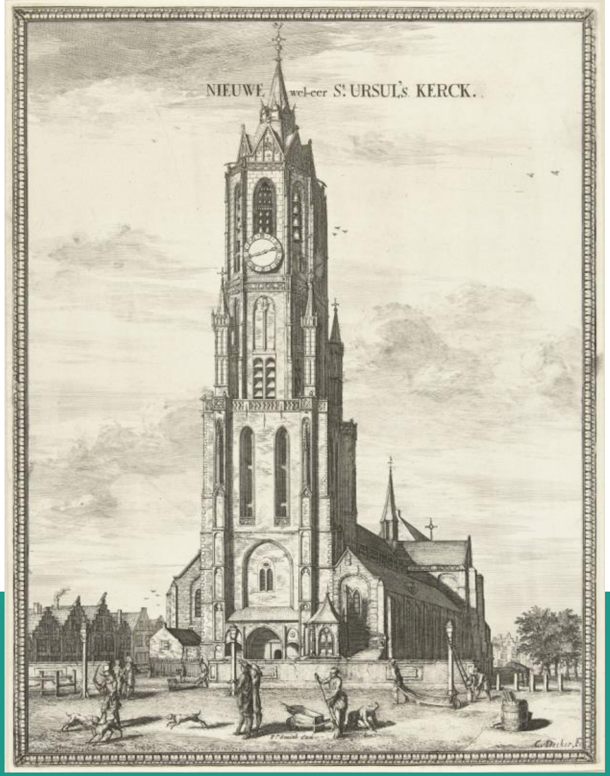
Simon Stevin & Jan Cornets de Groot in 1586:
Two balls with different weights, dropped from a height of 30 feet
[Coenraet Decker, Pieter Smith, Arnold Bon (1667)]

SOME HALLMARKS OF SCIENCE

- Generalization

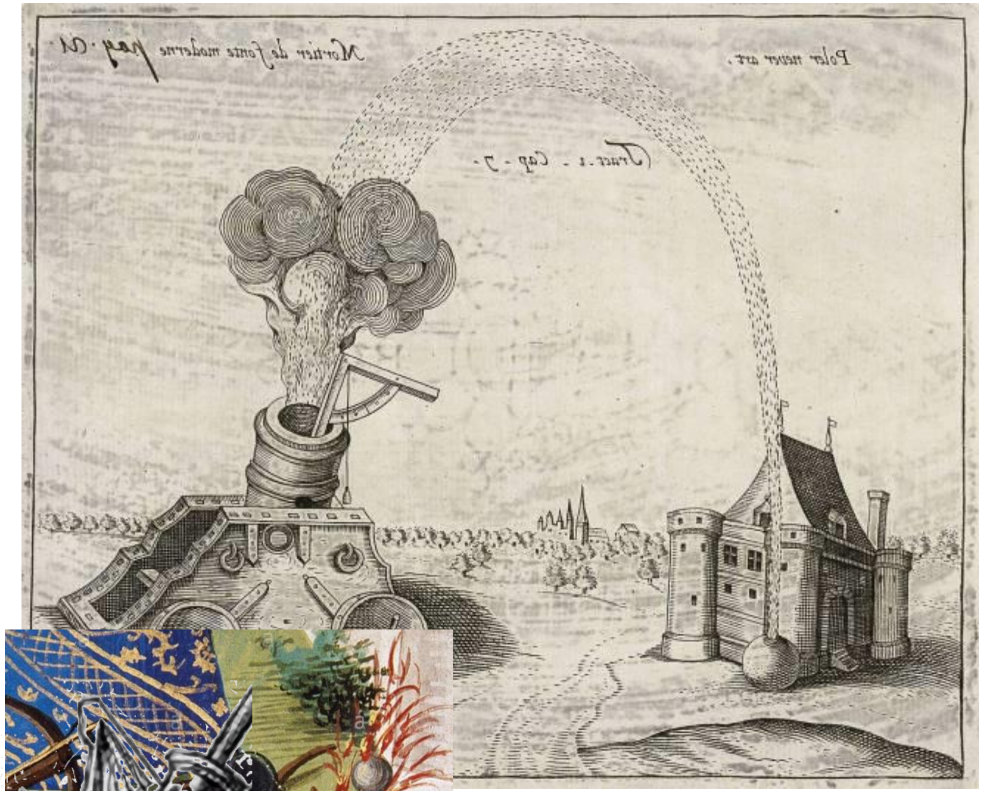
Derive findings that also apply to other experiments

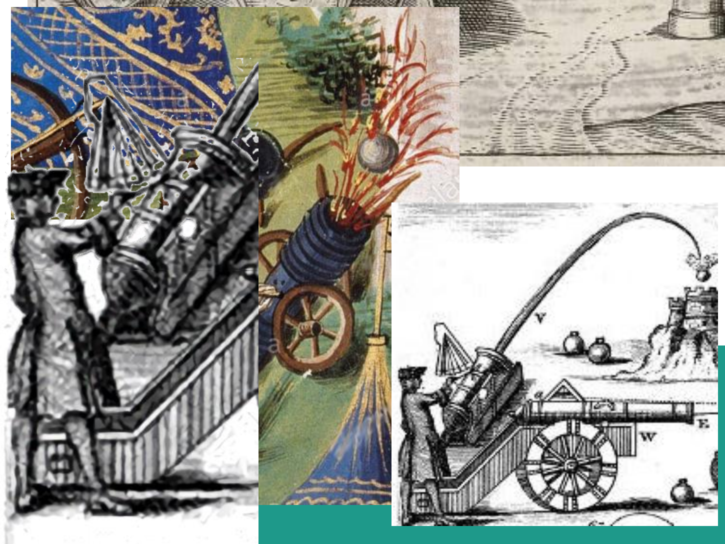
David Scott (taking part in NASA's Apollo 15 mission) reproducing on the moon an experiment demonstrating objects falling at the same speed. He released simultaneously a hammer and a feather from the same height: they landed at the same time.



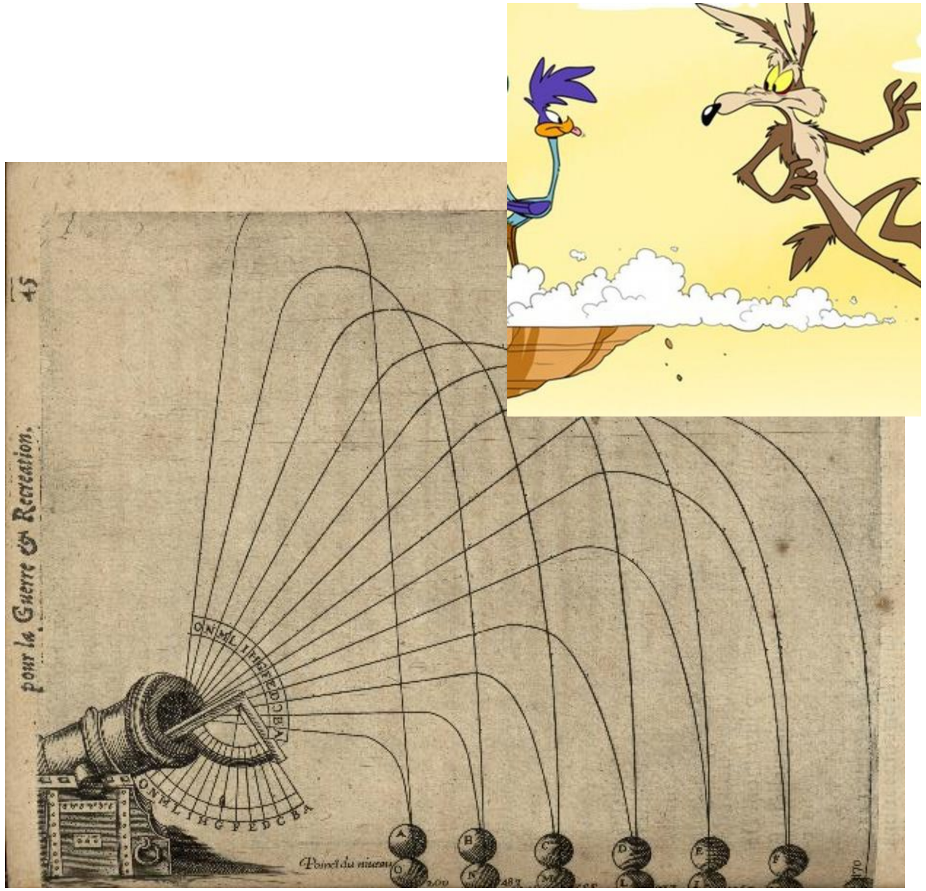
Simon Stevin & Jan Cornets de Groot in 1586:
Two balls with different weights, dropped from a height of 30 feet
[Coenraet Decker, Pieter Smith, Arnold Bon (1667)]

EXAMPLE OF MACHINE LEARNING APPLIED TO PROJECTILES





Mortier de fonte moderne, J. T. de Bry, 1613



La pyrotechnie de Hanzelet Lorrain, J. A. Hanzelet, 1630

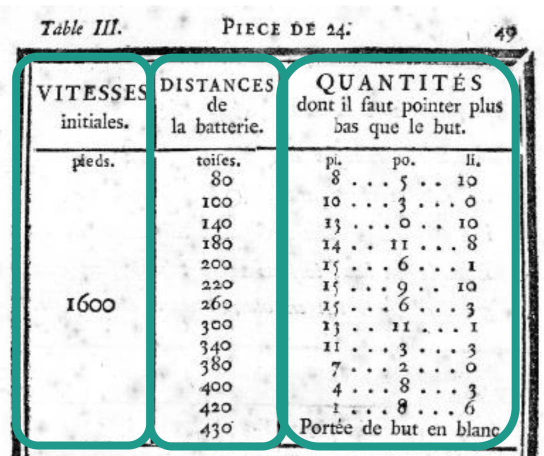


Table III. PIECE DE 24. 49

VITESSES initiales.	DISTANCES de la batterie.	QUANTITÉS dont il faut pointer plus bas que le but.		
pieds.	toises.	pi.	po.	li.
	80	8	5	10
	100	10	3	0
	140	13	0	10
	180	14	11	8
	200	15	6	1
	220	15	9	10
1600	260	15	6	3
	300	13	11	1
	340	11	3	3
	380	7	2	0
	400	4	8	3
	420	1	8	6
	430	Portée de but en blanc		

Initial speed | Range | ~ Angle

Tables du tir des canons et des obusiers, J. L.Lombard, 1787

EXAMPLE OF MACHINE LEARNING APPLIED TO PROJECTILES

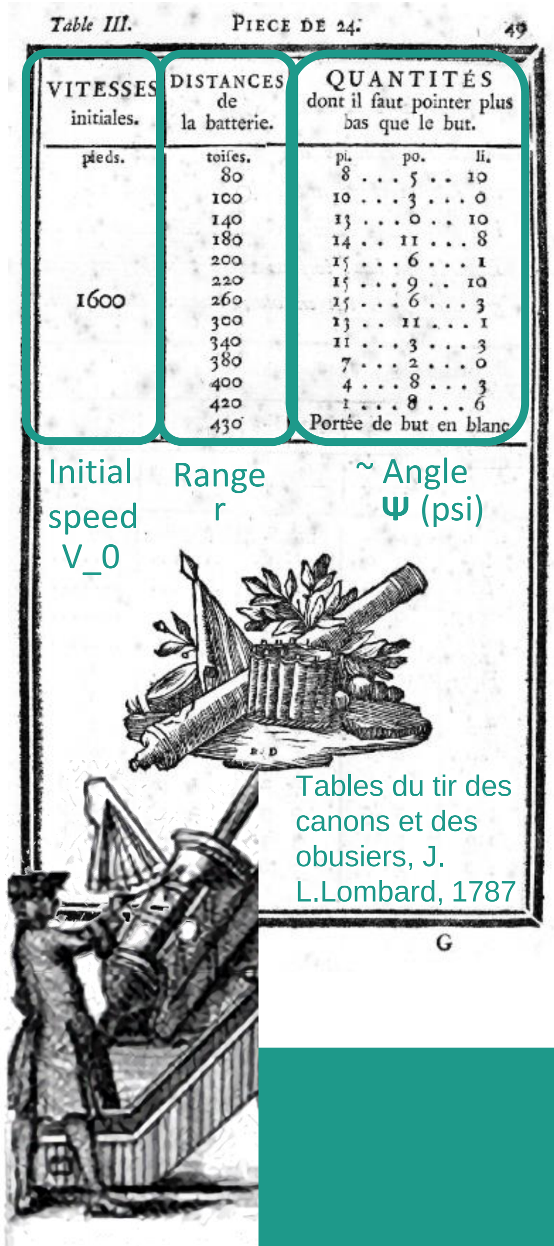


Table III. PIECE DE 24. 49

VITESSES initiales.	DISTANCES de la batterie.	QUANTITÉS dont il faut pointer plus bas que le but.
pieds.	toises.	pi. po. li.
1600	80	8 ... 5 .. 10
	100	10 ... 3 ... 0
	140	13 ... 0 .. 10
	180	14 .. 11 ... 8
	200	15 ... 6 ... 1
	220	15 ... 9 .. 10
	260	15 ... 6 ... 3
	300	13 .. 11 ... 1
	340	11 ... 3 ... 3
	380	7 ... 2 ... 0
	400	4 ... 8 ... 3
	420	1 ... 8 ... 6
	430	Portée de but en blanc

G

Tables du tir des canons et des obusiers, J. L.Lombard, 1787

- Take measurements
- Find a way to make predictions
- Assess generalizability of predictions (accuracy)

The choice of outcome(s) (and features) is fully part of the research design

DATASET

Features: Angle Ψ (psi), Initial speed, Mass, Radius, …
Labels (outcomes): Hit Target at 100 m
Observations (data points / samples)

	Angle Ψ (psi)	Initial speed	Mass	Radius	…	Hit Target at 100 m
Shot 1	60	20	3	0.10	…	No
Shot 2	45	20	3	0.10	…	Yes
Shot 3	45	20	5	0.12	…	No
Shot 4	35	20	4	0.08	…	Yes
…	…	…	…	…	…	…

CHOICE OF OUTCOME (AND FEATURES)

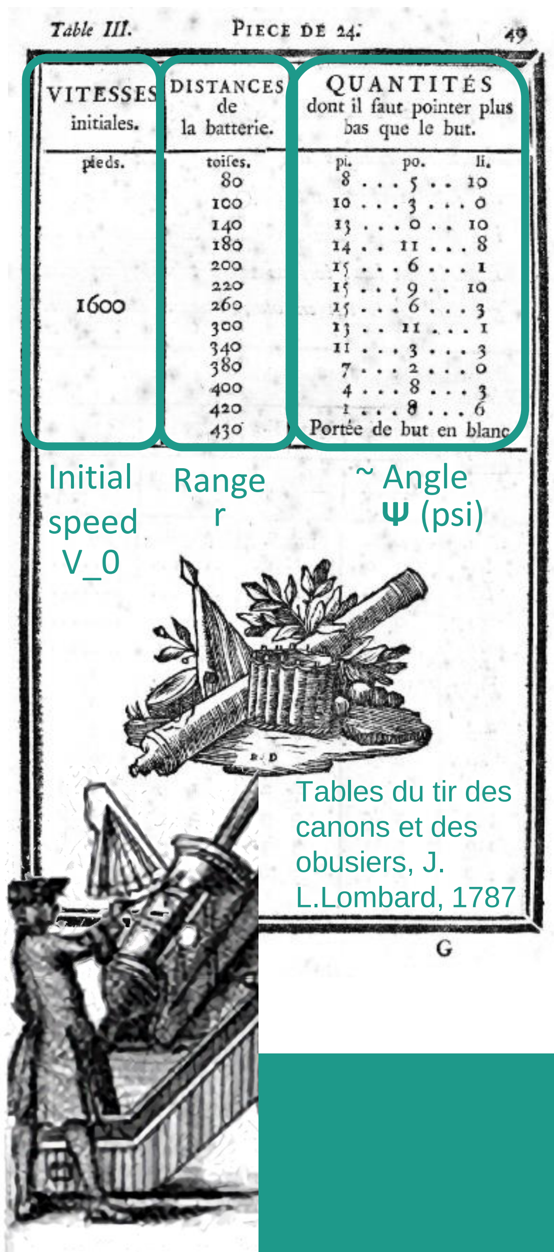


Table III. PIECE DE 24. 49

VITESSES initiales.	DISTANCES de la batterie.	QUANTITÉS dont il faut pointer plus bas que le but.
pieds.	toises.	pi. po. li.
	80	8 . . . 5 . . 10
	100	10 . . . 3 . . . 0
	140	13 . . . 0 . . 10
	180	14 . . 11 . . . 8
	200	15 . . . 6 . . . 1
	220	15 . . . 9 . . 10
1600	260	15 . . . 6 . . . 3
	300	13 . . 11 . . . 1
	340	11 . . . 3 . . . 3
	380	7 . . . 2 . . . 0
	400	4 . . . 8 . . . 3
	420	1 . . . 8 . . . 6
	430	Portée de but en blanc

Tables du tir des canons et des obusiers, J. L.Lombard, 1787

- ➢ Take measurements
- ➢ Find a way to make predictions
- ➢ Assess generalizability of predictions (accuracy)

The choice of outcome(s) (and features) is fully part of the research design

DATASET

Features | Labels (outcomes)

Observations (data points / samples)

	Age	Sex	ROI 1	ROI 2	…	Has disease
Subj 1	60	F	42.0	0.15	…	No
Subj 2	45	M	29.1	0.11	…	Yes
Subj 3	45	F	31.7	0.12	…	No
Subj 4	35	F	25.4	0.14	…	Yes
…	…	…	…	…	…	…

CHOICE OF OUTCOME (AND FEATURES)

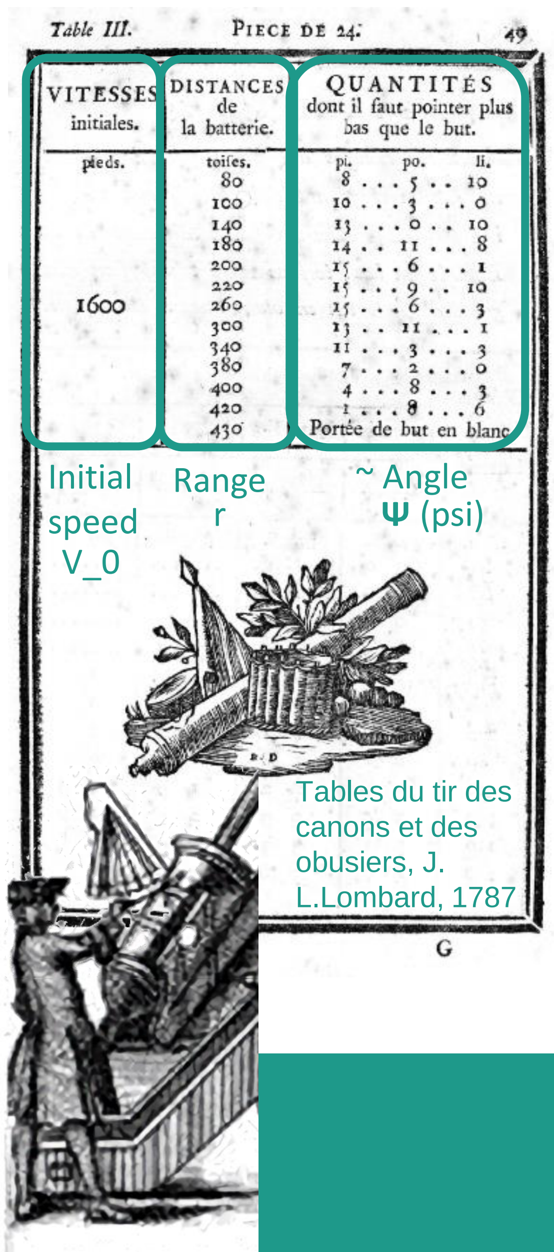


Table III. PIECE DE 24. 49

VITESSES initiales.	DISTANCES de la batterie.	QUANTITÉS dont il faut pointer plus bas que le but.		
pieds.	toises.	pi.	po.	li.
1600	80	8	5	10
	100	10	3	0
	140	13	0	10
	180	14	11	8
	200	15	6	1
	220	15	9	10
	260	15	6	3
	300	13	11	1
	340	11	3	3
	380	7	2	0
	400	4	8	3
	420	1	8	6
	430	Portée de but en blanc.		

G

Tables du tir des canons et des obusiers, J. L.Lombard, 1787

- ➢ Take measurements
- ➢ Find a way to make predictions
- ➢ Assess generalizability of predictions (accuracy)

The choice of outcome(s) (and features) is fully part of the research design

Binary diagnostic can impair understanding of disease:

- People with the same diagnostic can have different symptoms
- People with given symptoms can be likely to have an additional disorder
- Difficulty of clear diagnosis may cause to exclude patients from studies
- Criteria to be diagnose with a disorder can be arbitrary

➔ Choose biological, physiological, and behavioral dimensions as outcome

https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc/about-rdoc.shtml

EXAMPLE OF MACHINE LEARNING APPLIED TO PROJECTILES

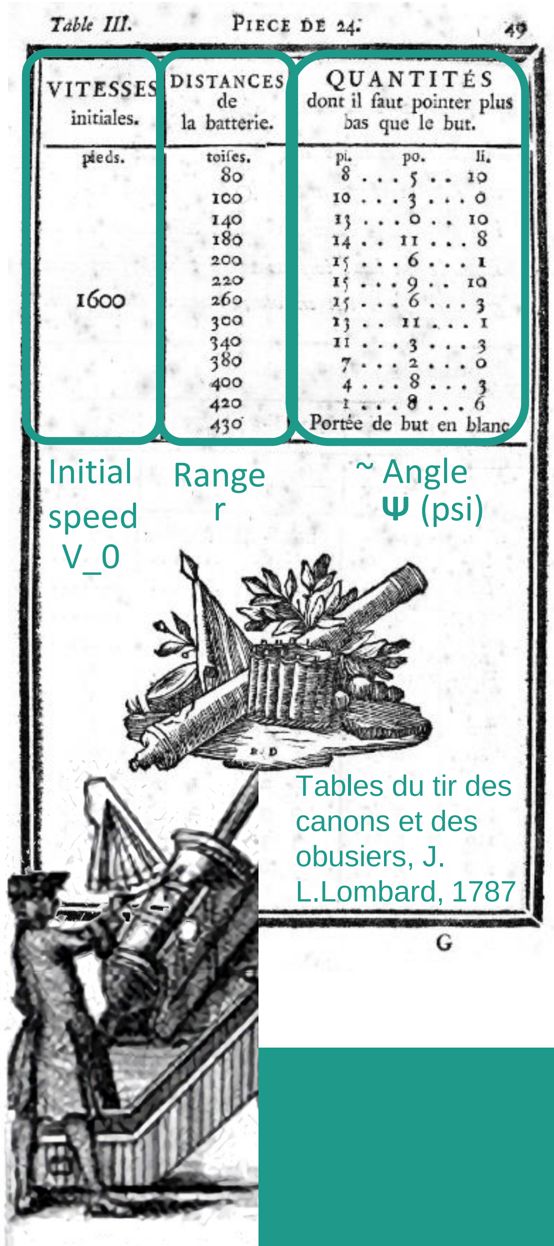


Table III. PIECE DE 24. 49

VITESSES initiales.	DISTANCES de la batterie.	QUANTITÉS dont il faut pointer plus bas que le but.		
pieds.	toises.	pi.	po.	li.
	80	8	5	10
	100	10	3	0
	140	13	0	10
	180	14	11	8
	200	15	6	1
	220	15	9	10
1600	260	15	6	3
	300	13	11	1
	340	11	3	3
	380	7	2	0
	400	4	8	3
	420	1	8	6
	430	Portée de but en blanc		

G

Tables du tir des canons et des obusiers, J. L.Lombard, 1787

- Take measurements
- Find a way to make predictions
- Assess generalizability of predictions (accuracy)

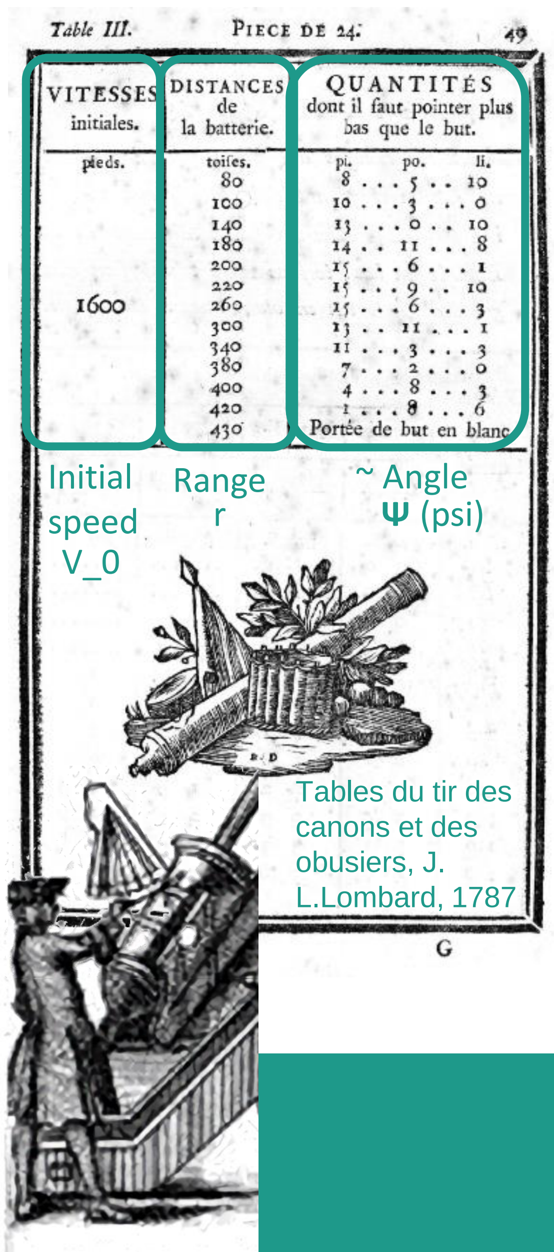
The choice of outcome(s) (and features) is fully part of the research design

DATASET

Observations (data points / samples)

	Features					Labels (outcomes)
	Angle Ψ (psi)	**Initial speed**	**Mass**	**Radius**	**…**	**Hit Target at 100 m**
Shot 1	60	20	3	0.10	…	No
Shot 2	45	20	3	0.10	…	Yes
Shot 3	45	20	5	0.12	…	No
Shot 4	35	20	4	0.08	…	Yes
…	…	…	…	…	…	…

EXAMPLE OF MACHINE LEARNING APPLIED TO PROJECTILES



- Take measurements
- Find a way to make predictions
- Assess generalizability of predictions (accuracy)

The choice of outcome(s) (and features) is fully part of the research design

DATASET

Features — Labels (outcomes)

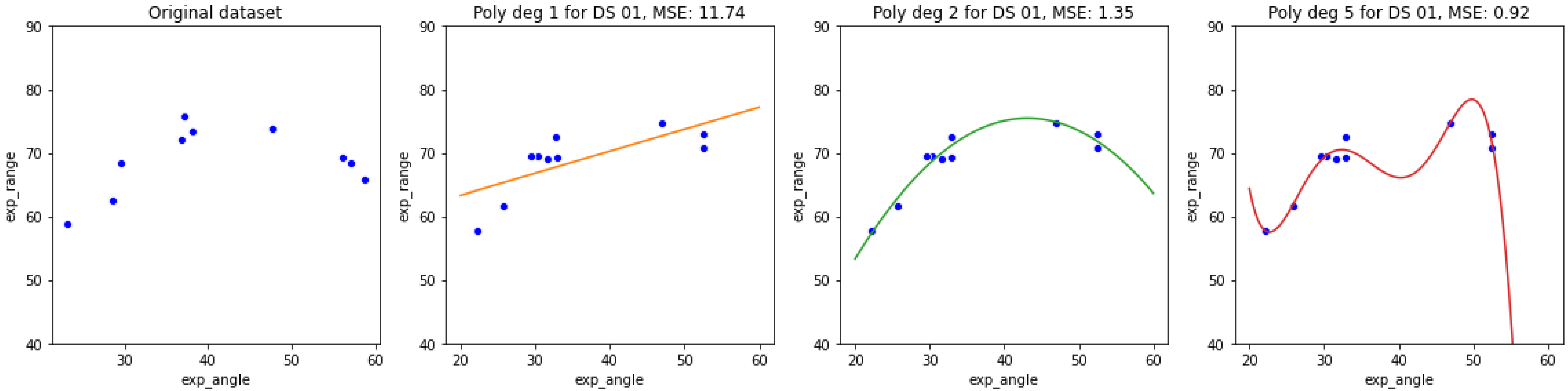
Observations (data points / samples)

	Angle Ψ (psi)	Initial speed	Mass	Radius	…	Range r
Shot 1	60	20	3	0.10	…	32.4
Shot 2	45	20	3	0.10	…	36.2
Shot 3	45	20	5	0.12	…	37.7
Shot 4	35	20	4	0.08	…	36.4
…	…	…	…	…	…	…

MAKING PREDICTIONS

"LEAST BAD" MODEL?

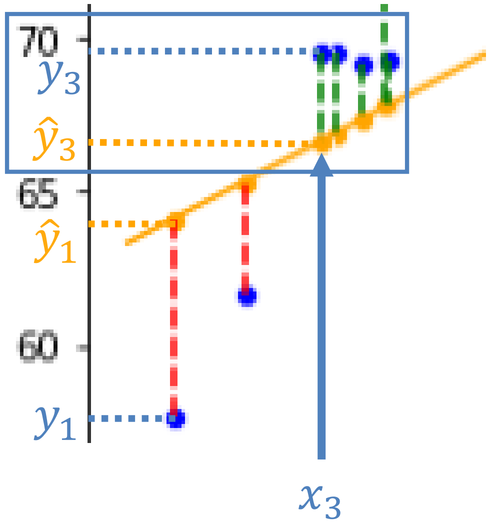
- Take measurements
- Find a way to make predictions
- Assess generalizability of predictions (accuracy)

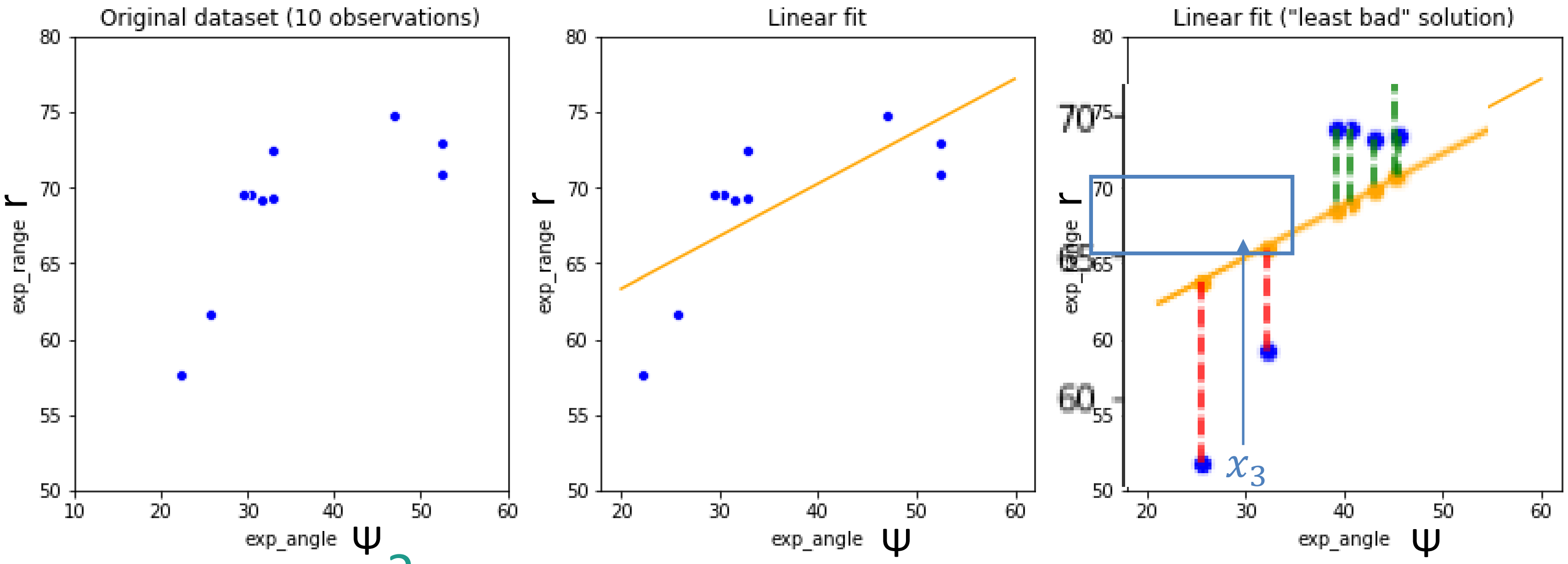


MAKING PREDICTIONS

- Take measurements
- Find a way to make predictions
- Assess generalizability of predictions

"LEAST BAD" MODEL?





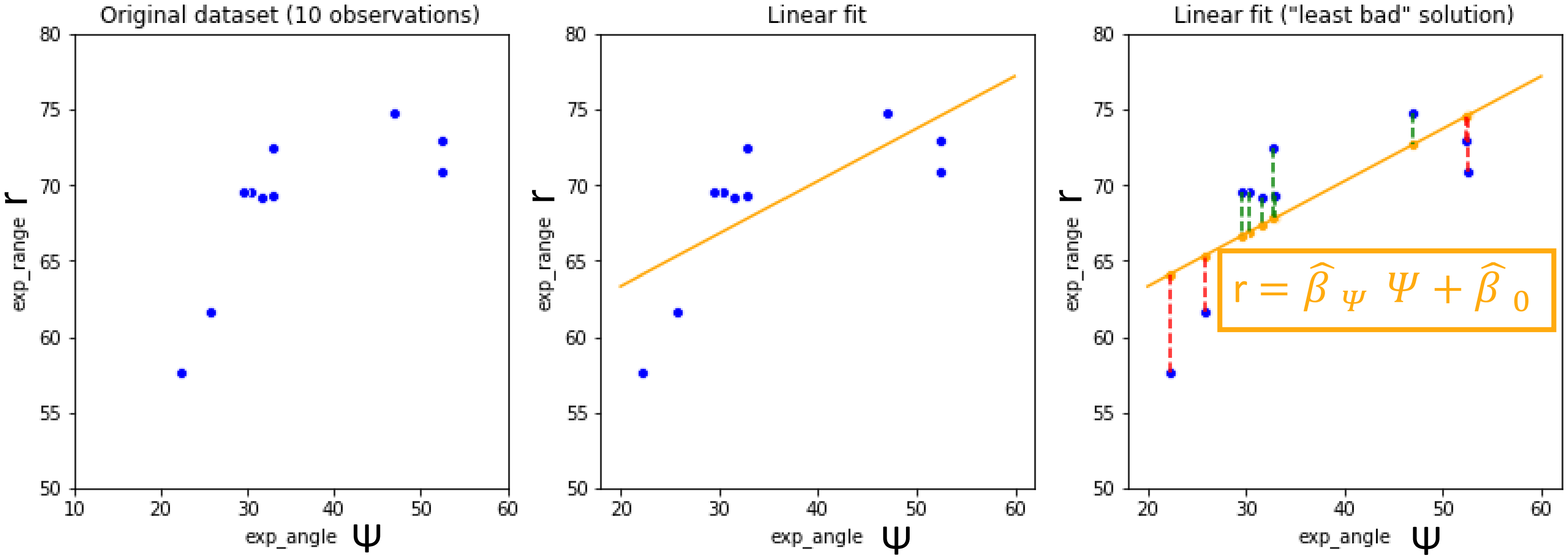
$$Error \stackrel{?}{=} (y_1 - \hat{y}_1) + (y_2 - \hat{y}_2) + (y_3 - \hat{y}_3) + \cdots + (y_N - \hat{y}_N)$$

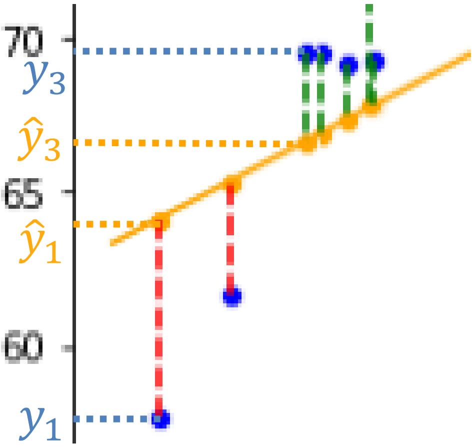
$$Squared\ Error = (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + \cdots + (y_N - \hat{y}_N)^2 = \sum_{i=0}^{N} (y_i - \hat{y}_i)^2$$

MAKING PREDICTIONS

- Take measurements
- Find a way to make predictions
- Assess generalizability of predictions

"LEAST BAD" MODEL?





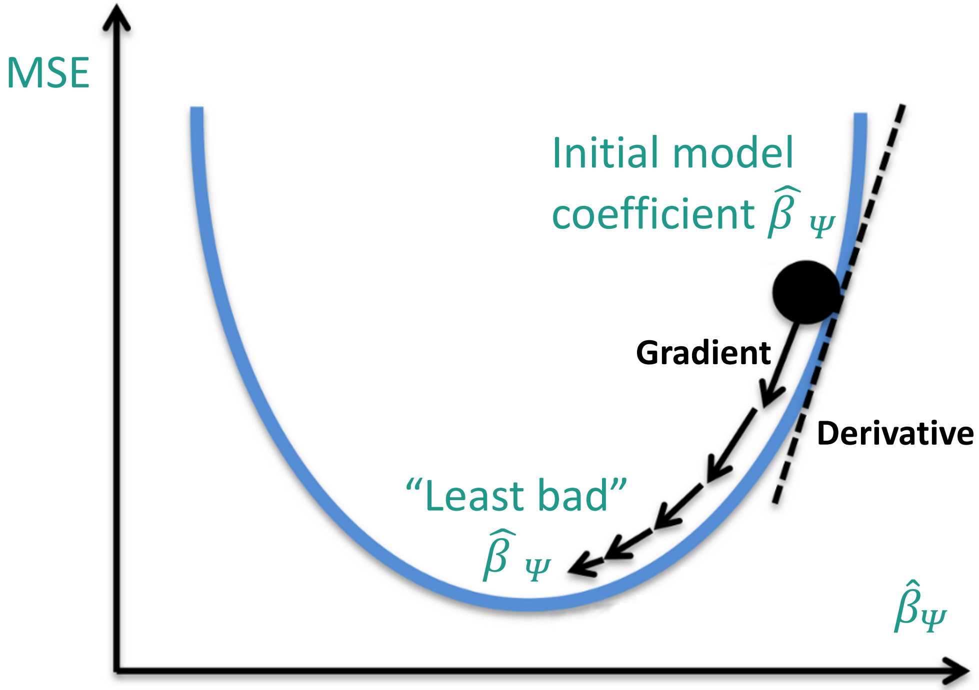
$$Mean\ Squared\ Error\ (MSE) = \frac{1}{N}\sum_{i=0}^{N}(y_i - \hat{y}_i)^2$$

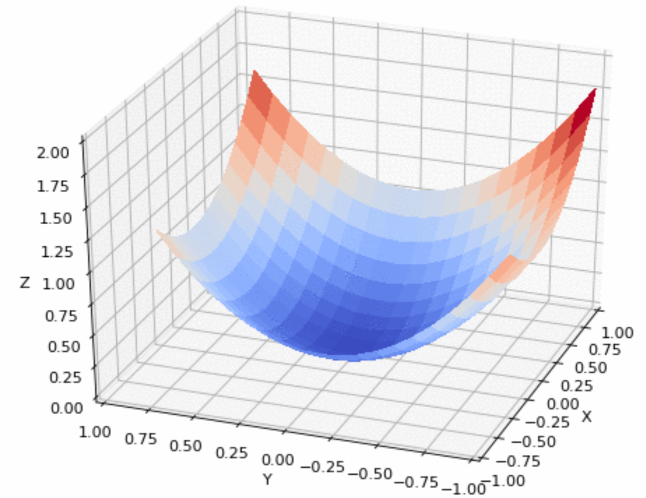
$$Error = (y_1 - \hat{y}_1) + (y_2 - \hat{y}_2) + (y_3 - \hat{y}_3) + \cdots + (y_N - \hat{y}_N)$$

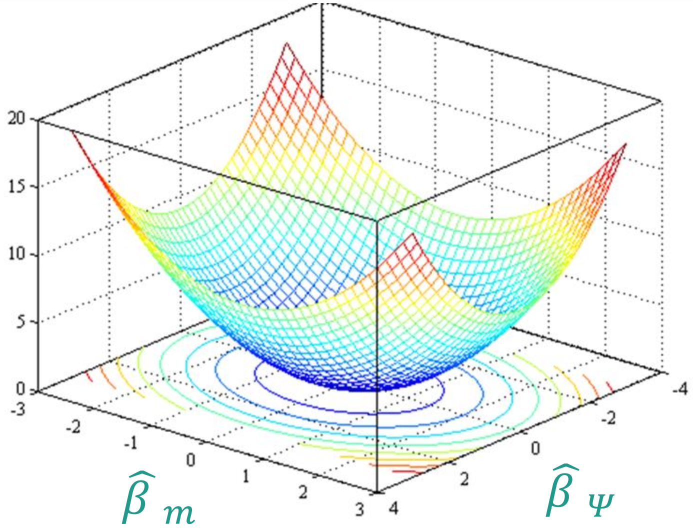
$$Squared\ Error = (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + \cdots + (y_N - \hat{y}_N)^2 = \sum_{i=0}^{N}(y_i - \hat{y}_i)^2$$

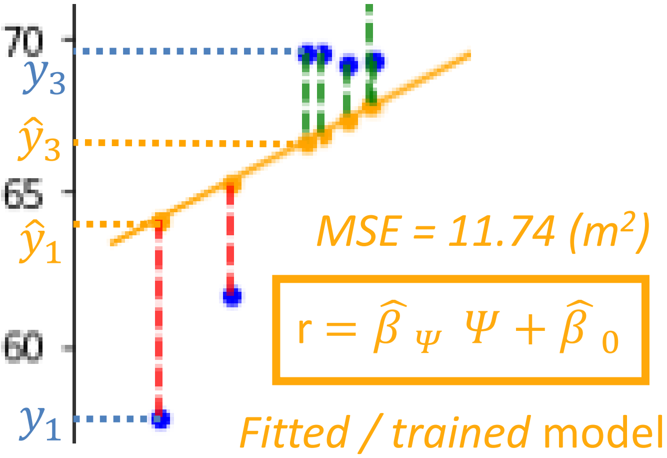
MAKING PREDICTIONS

- Take measurements
- Find a way to make predictions
- Assess generalizability of prediction







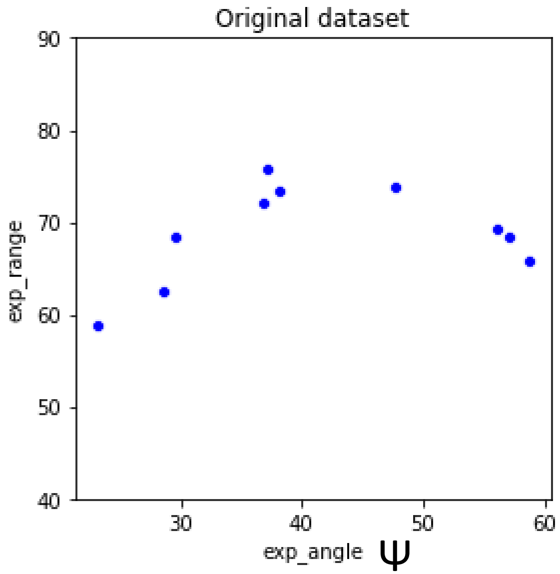


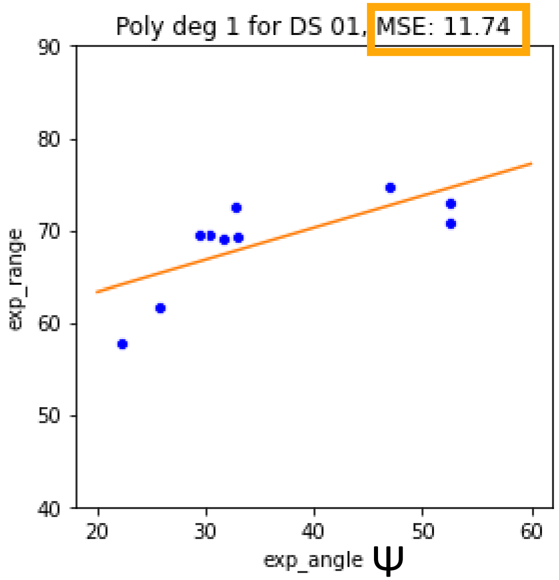
$$Mean\ Squared\ Error\ (MSE) = \frac{1}{N}\sum_{i=0}^{N}(y_i - \hat{y}_i)^2$$

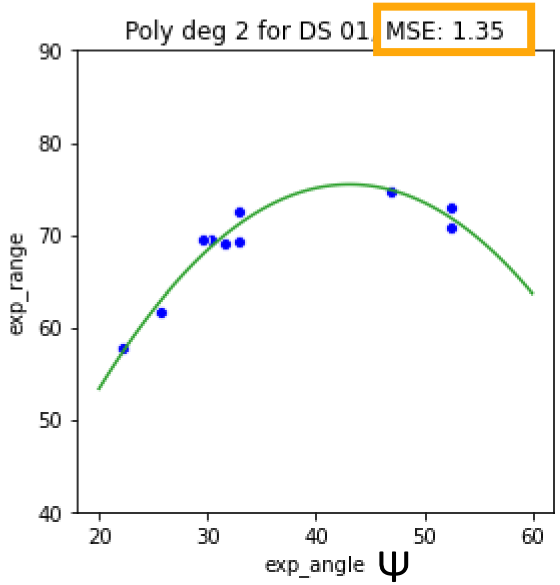
MAKING PREDICTIONS

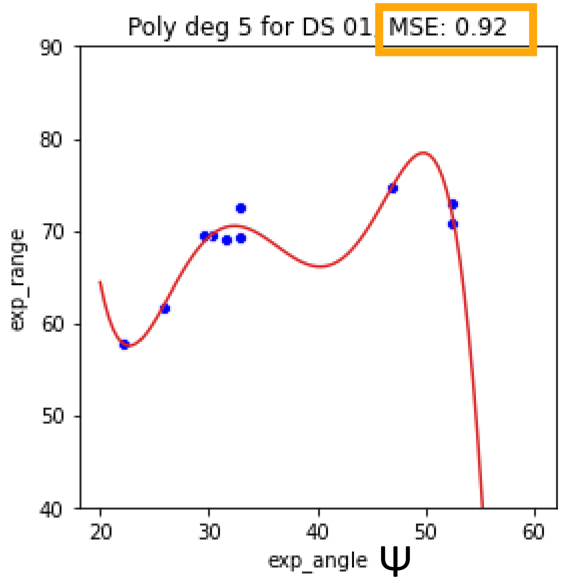
- Take measurements
- Find a way to make predictions
- Assess generalizability of predictions

"LEAST BAD" MODEL?









MSE = 11.74 (m^2)

$$\boxed{r = \widehat{\beta}_{\Psi}\,\Psi + \widehat{\beta}_0}$$

Fitted / trained model

$$r = \widehat{\beta}_{\Psi^2}\,\Psi^2 + \widehat{\beta}_{\Psi}\,\Psi + \widehat{\beta}_0$$

$$r = \widehat{\beta}_{\Psi^5}\,\Psi^5 + \widehat{\beta}_{\Psi^4}\,\Psi^4 + \cdots$$

MAKING PREDICTIONS

- Take measurements
- Find a way to make predictions
- Assess generalizability of predictions

.EAST BAD" MODEL?

Fitted model

$r = \hat{\beta}_{\Psi} \Psi + \hat{\beta}_0$

$X \rightarrow y_{pred}$

Test score (e.g. MSE)

ORIGINAL TRAINING DATASET

X — Features — Labels — y

Observations (data points / samples)

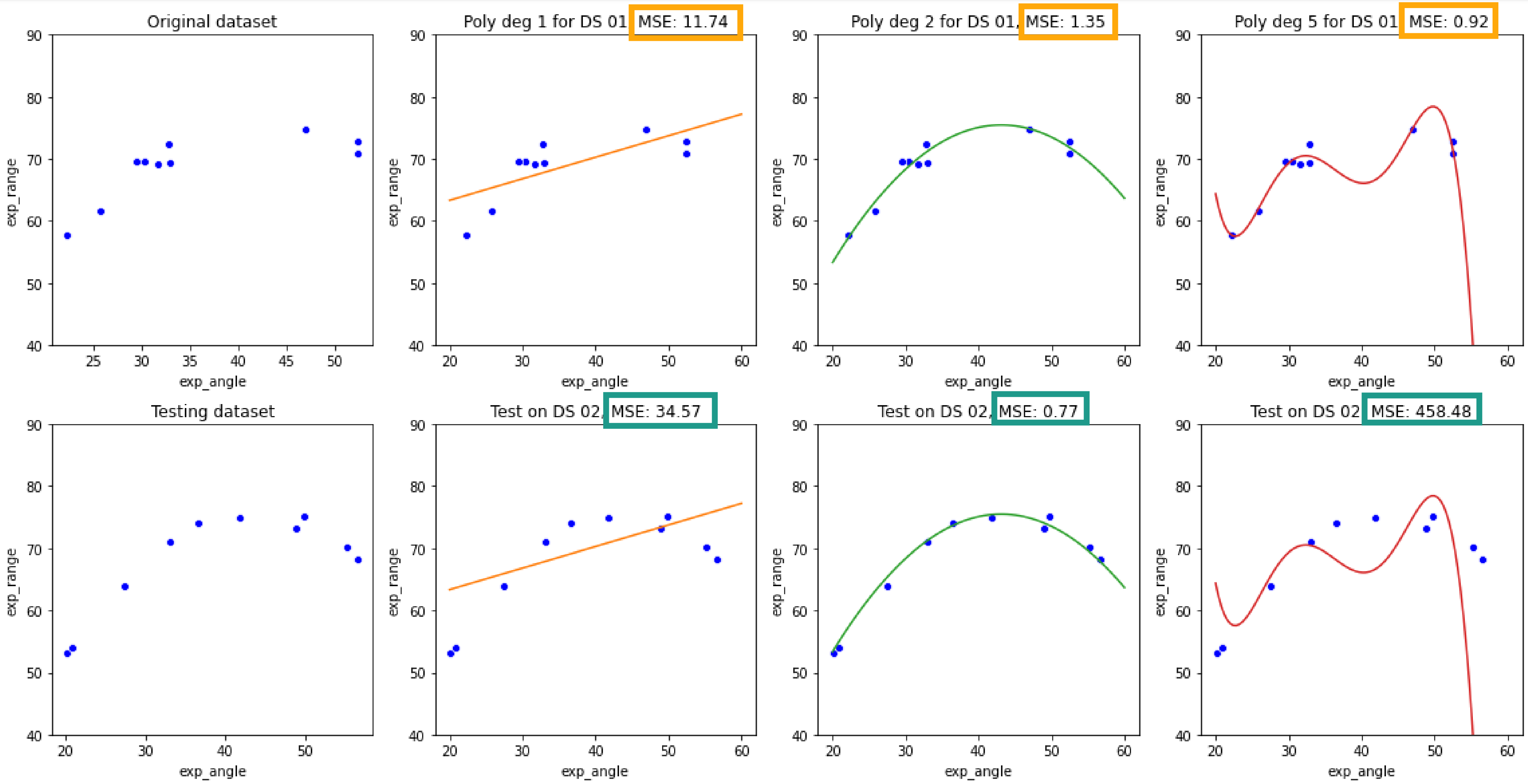
	Angle Ψ (psi)	Range r
Shot 1	60	32.4
Shot 2	45	36.2
Shot 3	45	37.7
…	…	…

NEW TESTING DATASET

Features — Labels — y

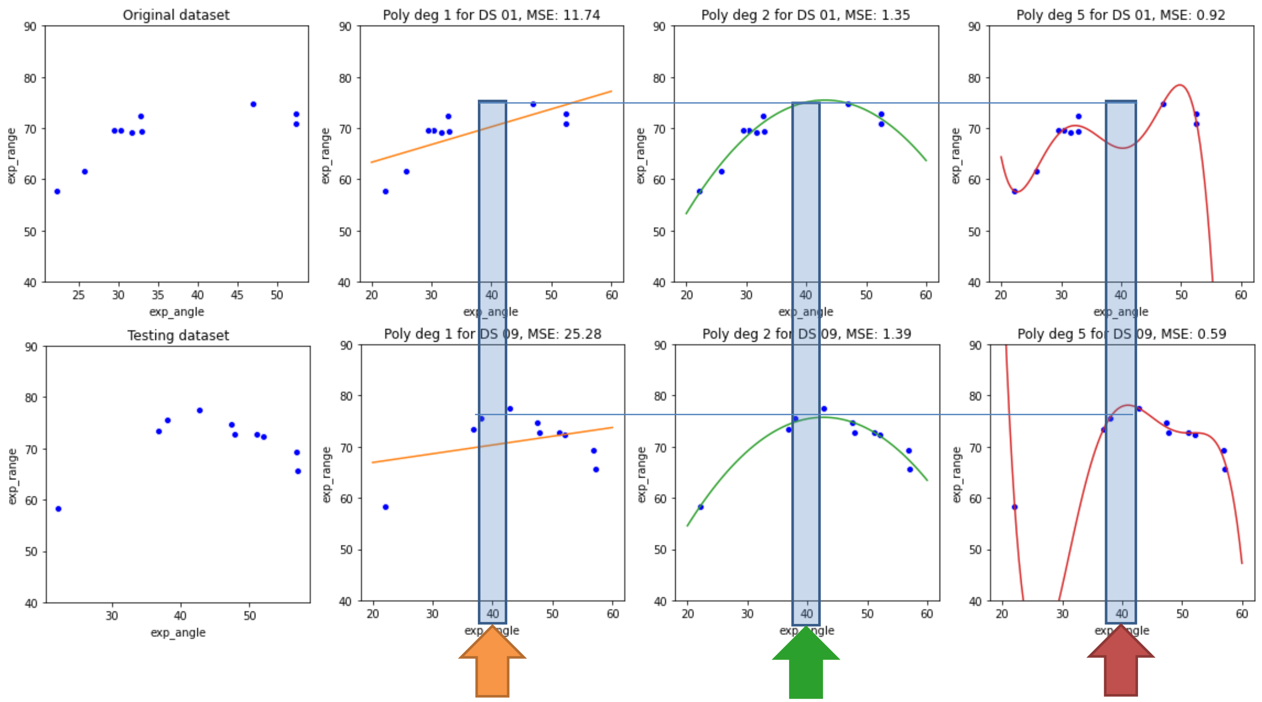
	Angle Ψ (psi)	Range r
Shot 1	37	33.1
Shot 2	24	29.6
Shot 3	54	41.4
…	…	…

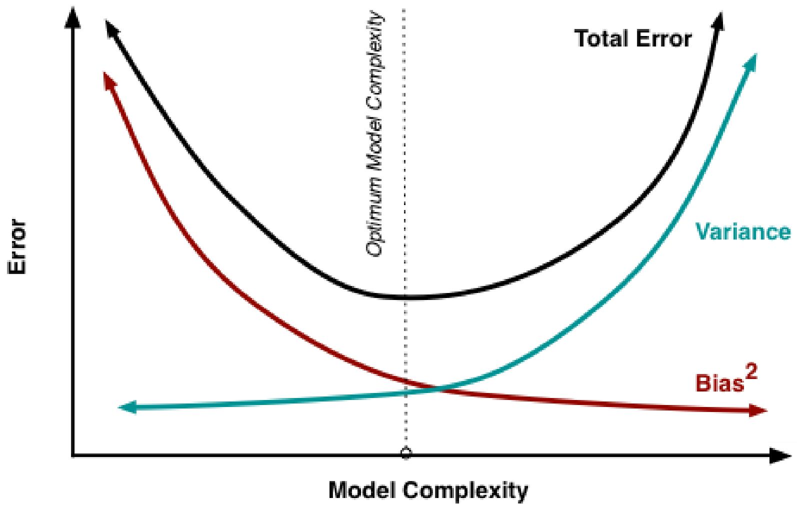
MAKING PREDICTIONS

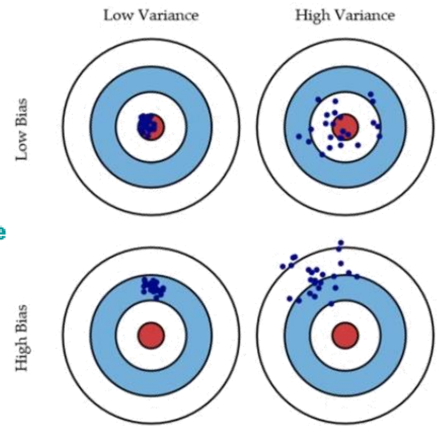


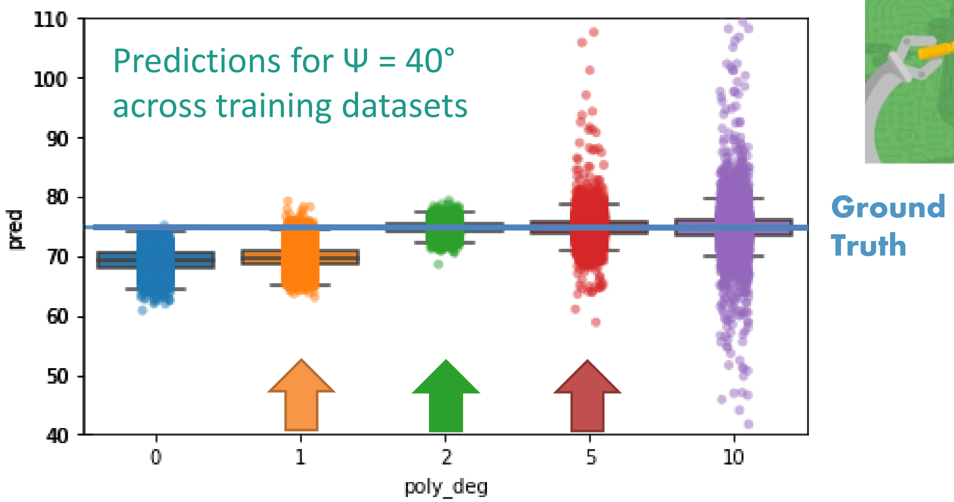
MAKING PREDICTIONS

- Take measurements
- Find a way to make predictions
- Assess generalizability of predictions







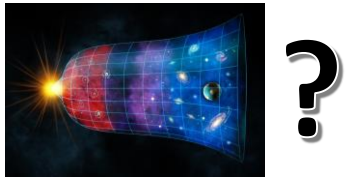


→ **Need a testing set: split your data in training and testing set**
Let's practice!

GENERALIZABILITY OF ML MODELS

- Take measurements
- Find a way to make predictions
- Assess generalizability of predictions

"Non-reproducible single occurrences are of no significance to science." *Karl Popper*



A hallmark of science is **generalization**:

→ derive findings that also apply to other experiments

➔ derive neuroimaging findings that apply to other population samples

MAKING PREDICTIONS

- Take measurements
- Find a way to make predictions
- Assess generalizability of predictions

STEP 1

Preprocessing

STEP 2

Train / Test split

Training set

Test set

Model selection

STEP 3

~~Model selection ?~~

Training set

(Sub) training set

Validation set

kaggle

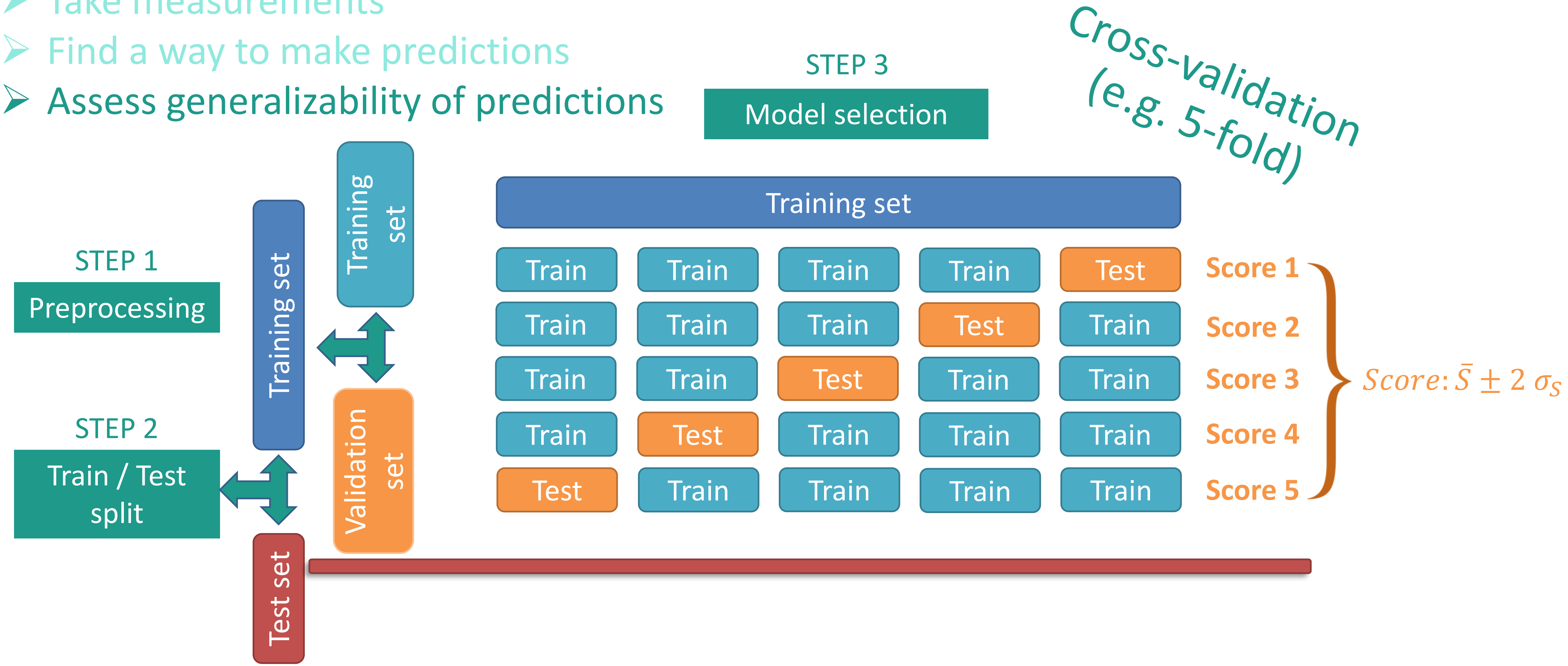
DREAM CHALLENGES

powered by Sage Bionetworks

MICCAI

GENERALIZABILITY OF ML MODELS

- Take measurements
- Find a way to make predictions
- Assess generalizability of predictions

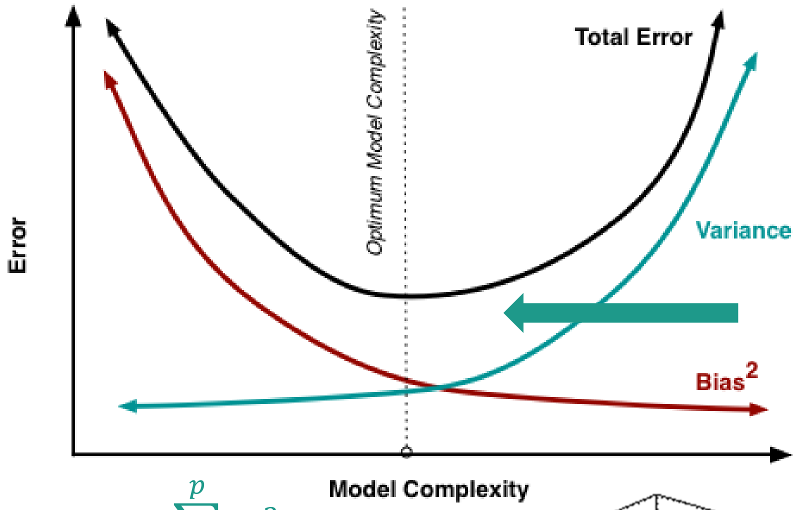


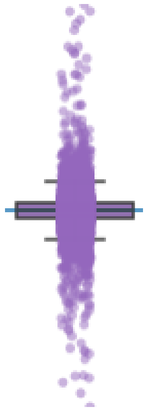
→ **Let's practice !**

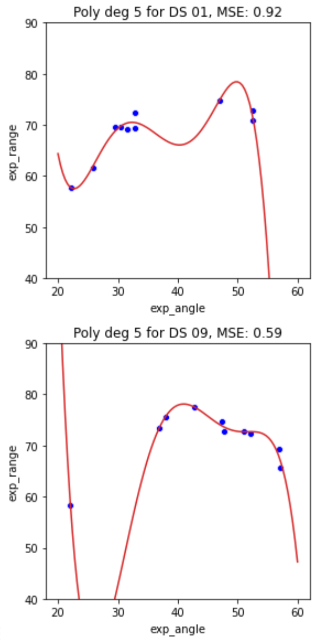
REGRESSION MODELS WITH REGULARIZATION

- How to prevent overfitting?
  - Feature selection
  - Regularization (in linear regression), forcing model parameters (β coefficients) to be small or zero

$\boxed{\lambda \text{ is an hyper-parameter (not fitted)}}$

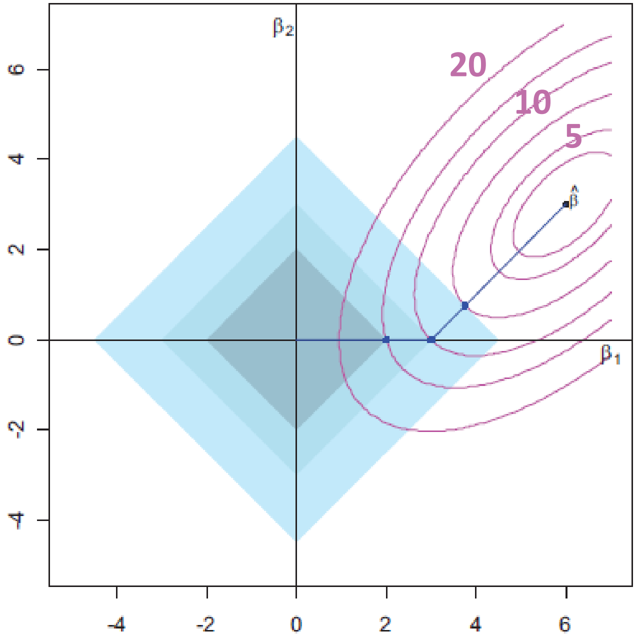


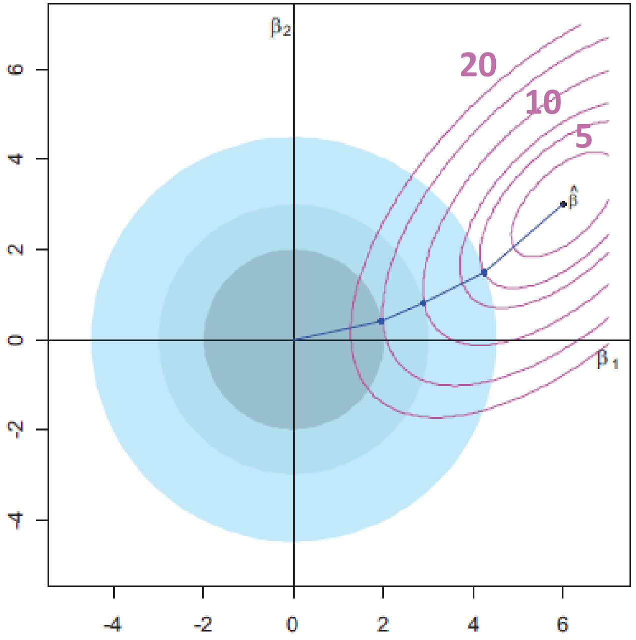


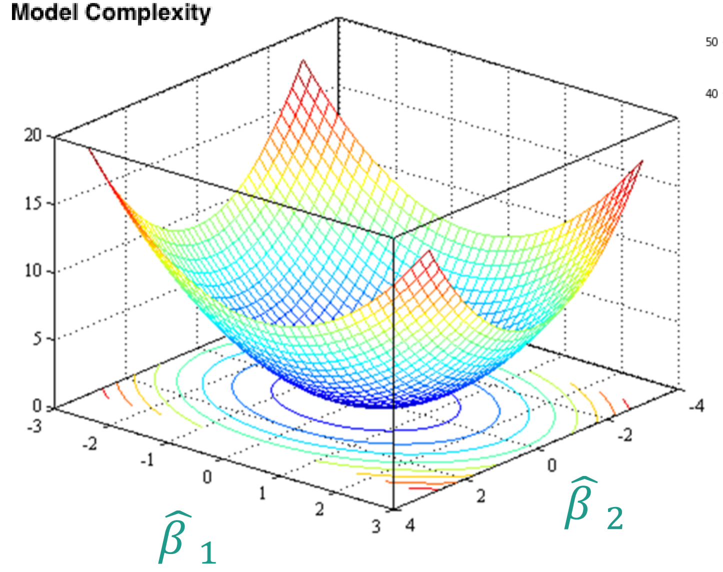


$$L_{lasso} = \sum_{i=0}^{n} (y_i - \hat{\beta} \cdot \boldsymbol{x}_i)^2 + \lambda \sum_{i=0}^{p} |\hat{\beta}_j|$$

$$L_{ridge} = \sum_{i=0}^{n} (y_i - \hat{\beta} \cdot \boldsymbol{x}_i)^2 + \lambda \sum_{i=0}^{p} \hat{\beta}_j^{\,2}$$







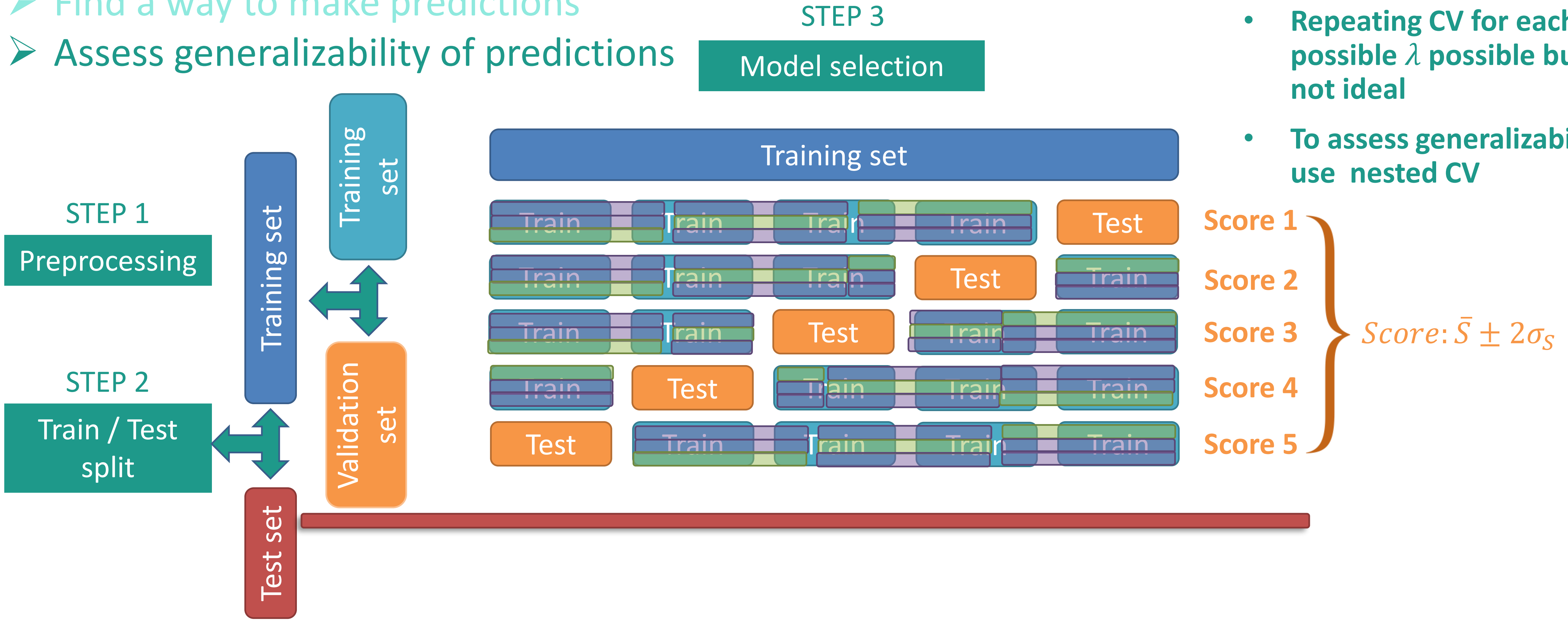
→ **Need to find best hyper-parameter**

REGRESSION MODEL WITH REGULARIZATION

- Take measurements
- Find a way to make predictions
- Assess generalizability of predictions

→ **How to find best hyper-parameter λ (e.g. among λ_1, λ_2 and λ_3)?**

- **Repeating CV for each possible λ possible but not ideal**
- **To assess generalizability, use nested CV**



→ Let's have a look at a typical workflow on real neuroimaging data

CLASSIFICATION & EVALUATION

- Classification scores

		Actual	
		Positive	Negative
Predicted	Positive	**True positive**	**False negative**
	Negative	**False positive**	**True negative**

Accuracy = (TP + TN) / (TP + FP + FN + TN)

Sensitivity (TPR) = TP / (TP+FN)

False Positive Rate (FPR) = FP / (FP+TN)

Specificity = TN / (FP+TN)

Precision (PPV) = TP / (TP+FP)

→ **Depends on which probability threshold you choose to define positive cases**
p=0.5? p=0.95?

The choice of outcome(s) (and features) is fully part of the research design

DATASET

Features — Labels (outcomes)

Observations (data points / samples)

	Age	Sex	ROI 1	…	Has disease
Subj 1	60	F	42.0	…	No
Subj 2	45	M	29.1	…	Yes
Subj 3	45	F	31.7	…	No
Subj 4	35	F	25.4	…	Yes
…	…	…	…	…	…

CLASSIFICATION & EVALUATION

- Classification scores

		Actual	
		Positive	Negative
Predicted	Positive	True positive	False negative
	Negative	False positive	True negative

Accuracy = (TP + TN) / (TP + FP + FN + TN)

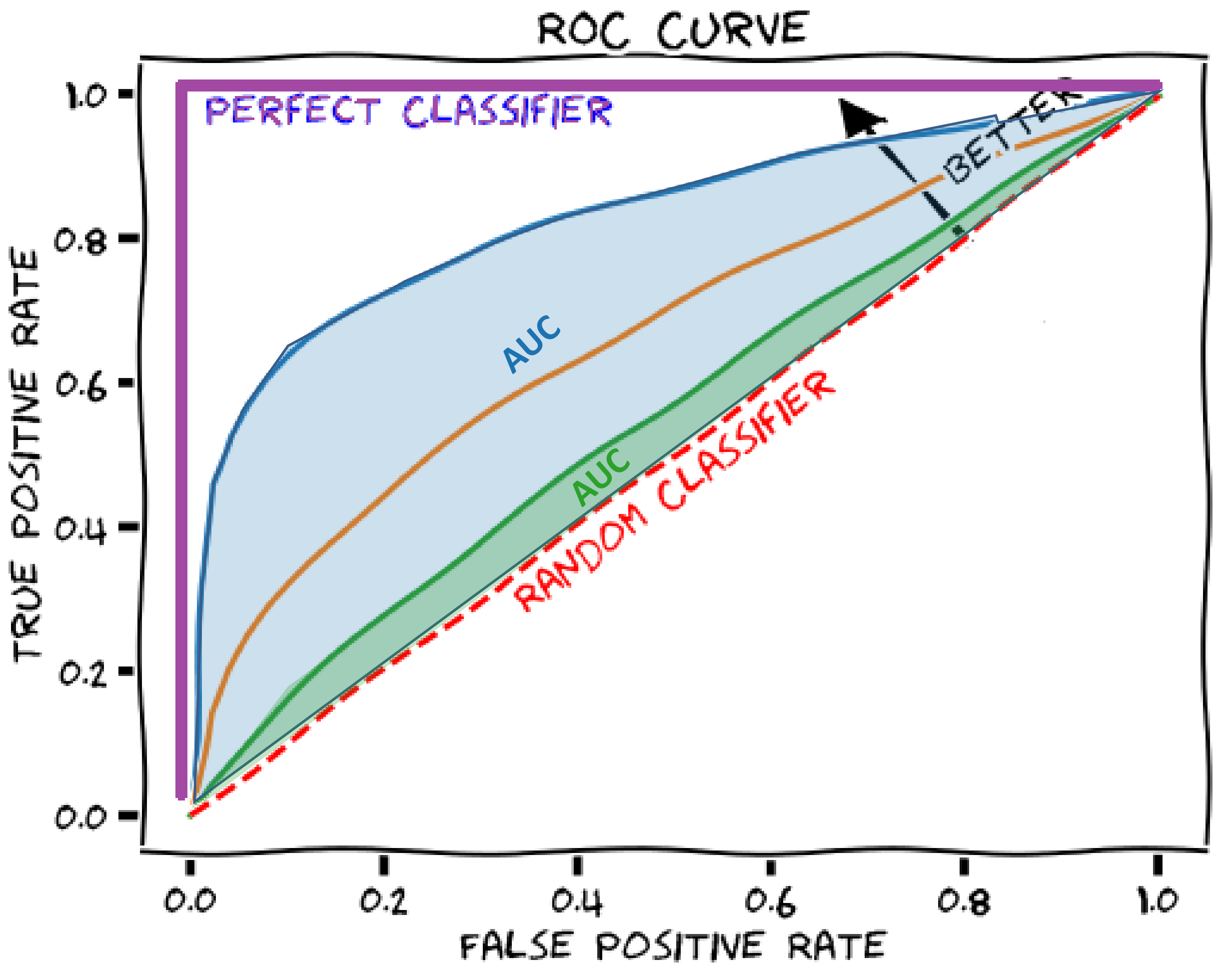
Sensitivity (TPR) = TP / (TP+FN)

False Positive Rate (FPR) = FP / (FP+TN)

Specificity = TN / (FP+TN)

Precision (PPV) = TP / (TP+FP)

→ **Depends on which probability threshold you choose to define positive cases**
p=0.5? p=0.95?



CLASSIFICATION EXAMPLE: SVM

$$\text{margin} = \frac{1}{\|w\|}$$

$$\underset{\mathbf{w},b}{\text{minimize}} \quad \frac{1}{2}||\mathbf{w}||^2$$

$$\text{subject to:} \quad y_i(\mathbf{w}^\mathsf{T}\mathbf{x}_i + b) \geq 1 \quad i = 1, \ldots, n\,.$$

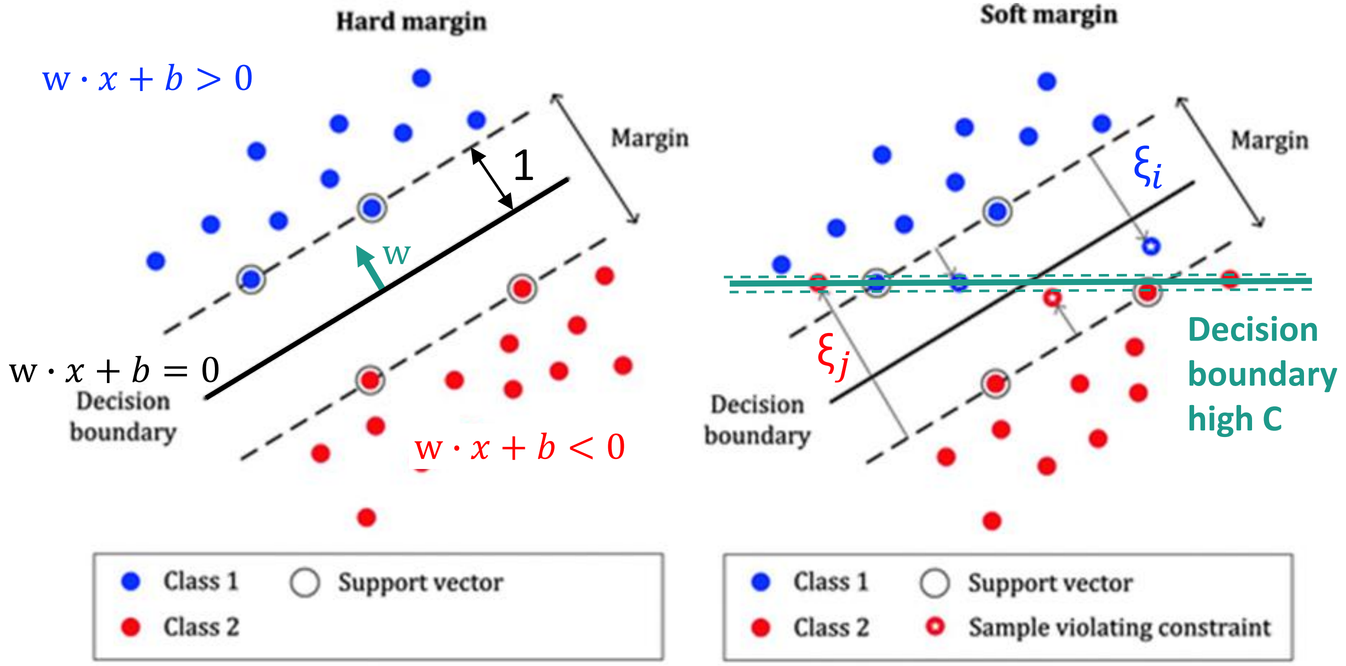
$$\underset{\mathbf{w},b}{\text{minimize}} \quad \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{n} \xi_i$$

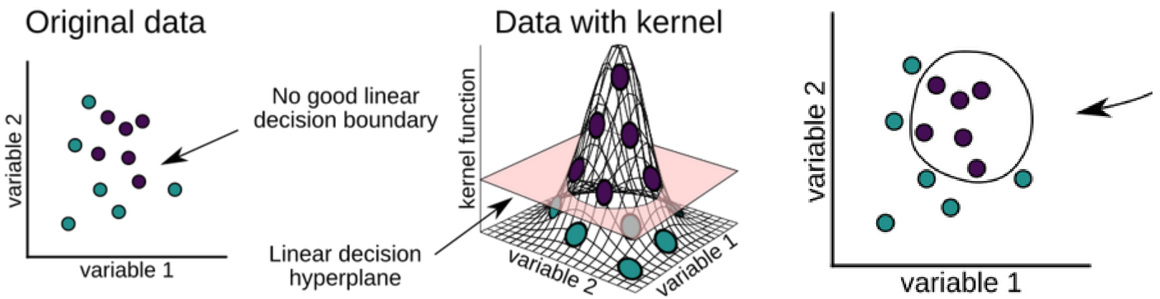
$$\text{subject to:} \quad y_i(\mathbf{w}^\mathsf{T}\mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0.$$

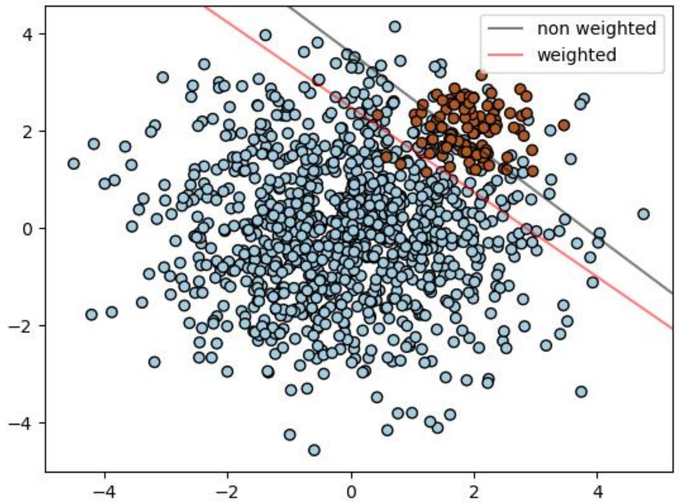
$$C\sum_{i=1}^{n}\xi_i \longrightarrow C_+\sum_{i\in I_+}\xi_i + C_-\sum_{i\in I_-}\xi_i$$

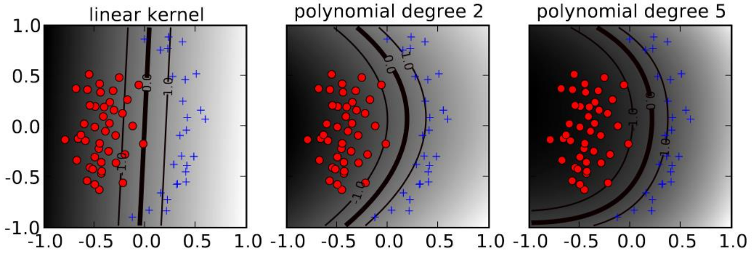
$$C_+ n_+ = C_- n_-$$

$$\frac{C_+}{C_-} = \frac{n_-}{n_+}.$$







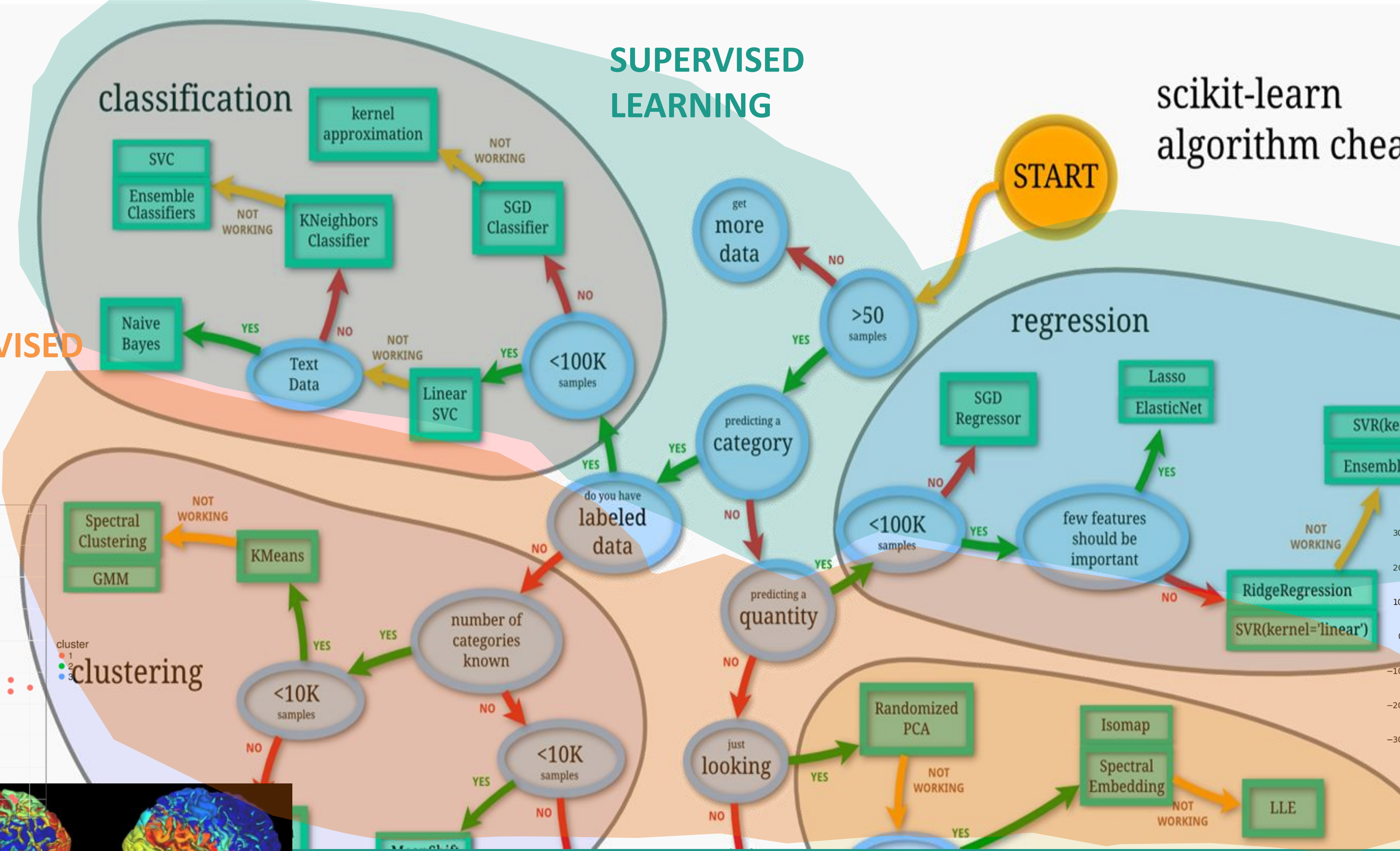


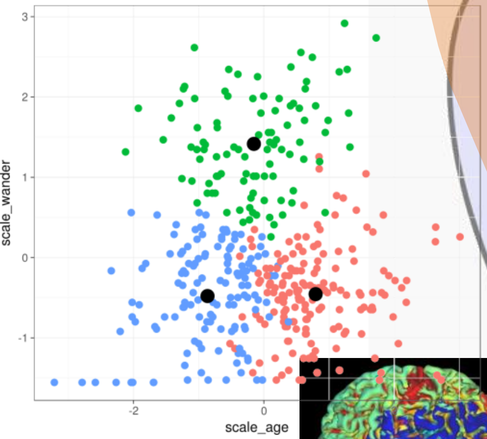
→ nilearn's Haxby dataset tutorial

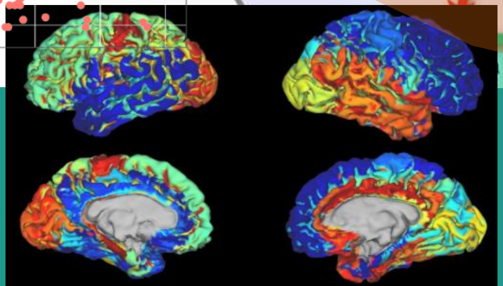
MACHINE LEARNING ESTIMATORS

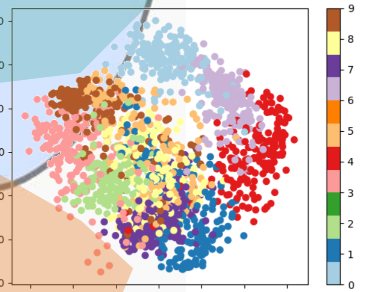
SUPERVISED LEARNING

UNSUPERVISED LEARNING









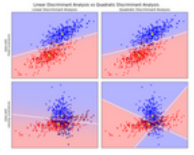
THANK YOU FOR YOUR ATTENTION!

Classification

General examples about classification algorithms.

Normal and Shrinkage Linear Discriminant Analysis for classification

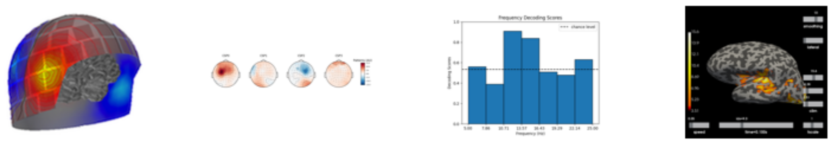
Re dig



Linear and Quadratic Discriminant Analysis with covariance ellipsoid

Machine Learning (Decoding, Encoding, and MVPA)

Decoding, encoding, and general machine learning examples.



Representational Similarity Analysis

Motor imagery decoding from EEG data using the Common Spatial Pattern (CSP)

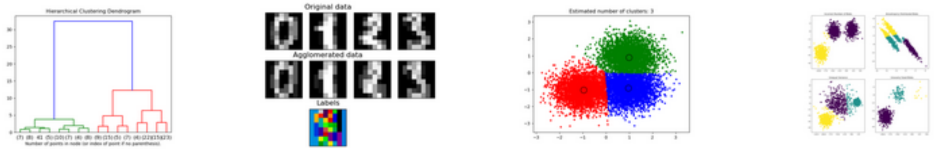
Decoding in time-frequency space data using the Common Spatial Pattern (CSP)

Decoding source space data

MEG + EEG ANALYSIS & VISUALIZATION

Clustering

Examples concerning the `sklearn.cluster` module.



Plot Hierarchical Clustering Dendrogram

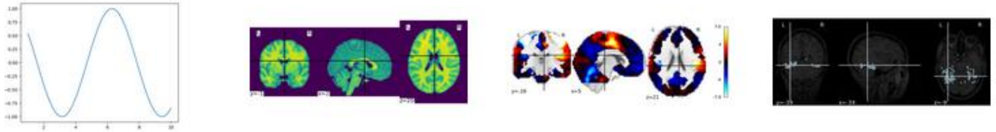
Feature agglomeration

A demo of the mean-shift clustering algorithm

Demonstration of k-means assumptions

8.1. Tutorial examples

Introductory examples that teach how to use nilearn.



Basic numerics and plotting with Python

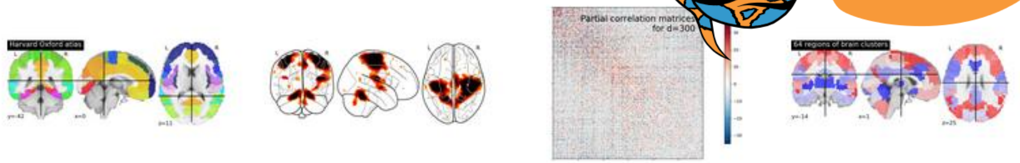
Basic nilearn example: manipulating and looking at data

3D and 4D niimgs: handling and visualizing

A introduction tutorial to fMRI decoding

8.2. Visualization of brain images

See Plotting brain images for more details.

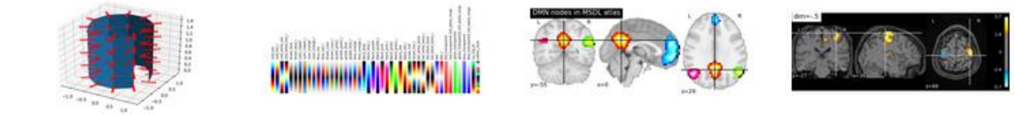


Basic Atlas plotting

Glass brain plotting in nilearn

Visualizing Megatrawls Network Matrices from Human Connectome Project

Visualizing multiscale functional brain parcellations



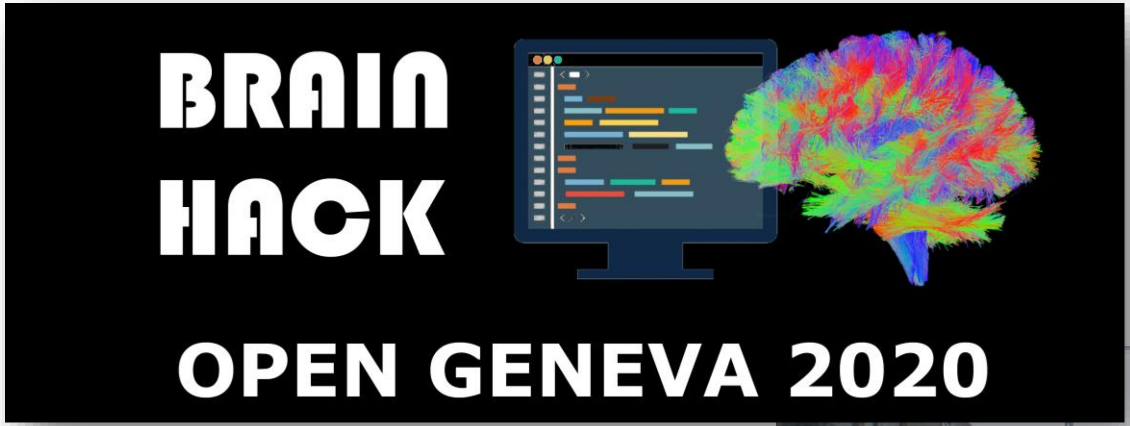
Technical point: Illustration of the volume to surface sampling

Matplotlib colormaps in Nilearn

Visualizing a probablistic atlas: the default mode in the MSDL atlas

Controling the contrast of the background when plotting

THANK YOU FOR YOUR ATTENTION!



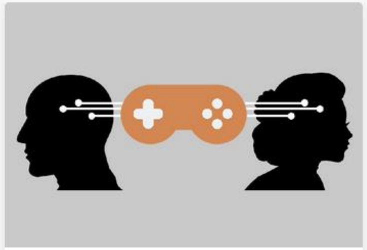
Autumn 2020

@brainhackgeneva

WWW.BRAINHACK.CH

Reconstruct visual stimuli from EEG data

BY KONSTANTINOS SAMARAS-TSAKIRIS



Cooperative Neurofeedback

BY LOUIS ALBERT AND SIXTO ALCOBA BANQUERI

PlasticBrain

BY MANIK BHATTACHARJEE