Imperial College London

Introduction to Machine Learning

Valentina Giunchiglia

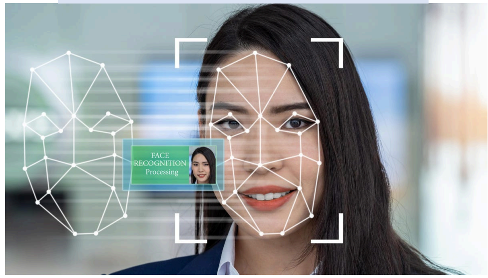
v.giunchiglia20@imperial.ac.uk

What is Machine Learning?

"Machine Learning: a computer is able to learn from experience without being specifically programmed"

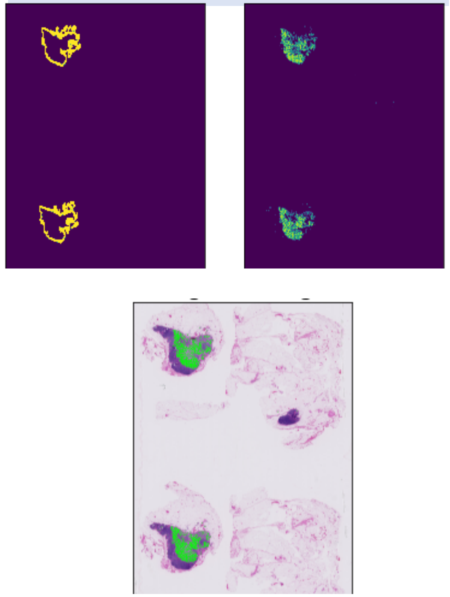
Self-drinving cars

Face recognition

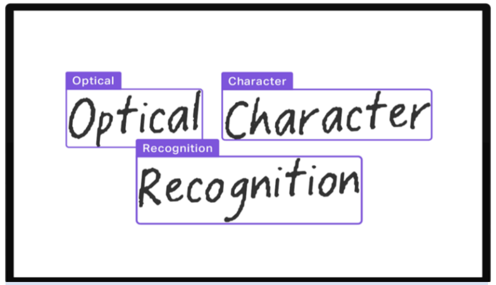


Content ads

Biomedical Research



Speech recognition

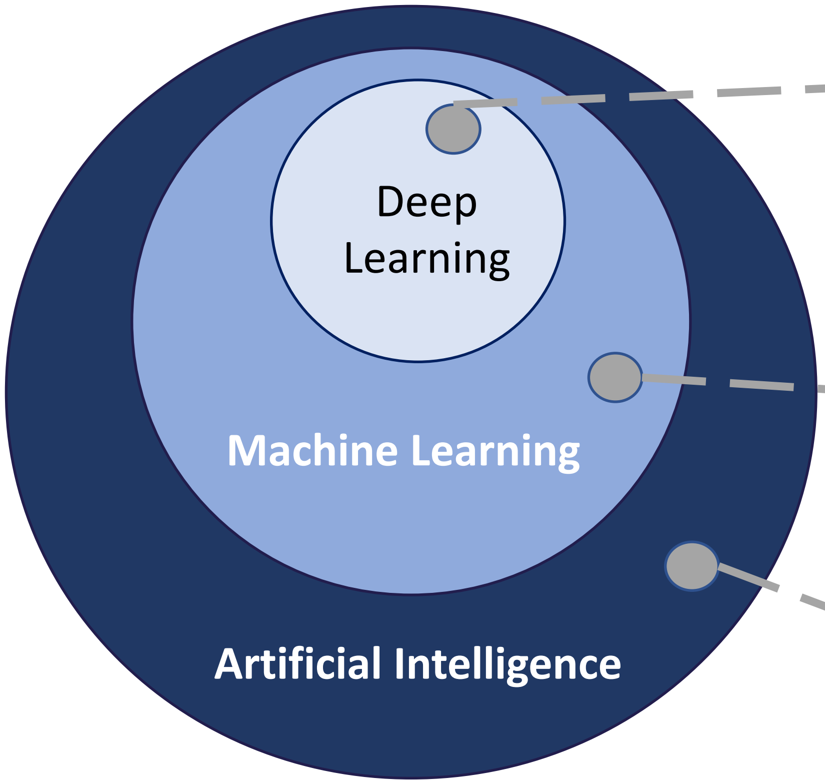


Character Recognition

Typing prediction

Imperial College London

What is Machine Learning?



Subset of machine learning that uses artifical neural networks — Stream 1

Application of AI that allows a machine to learn and improve from experience — Today

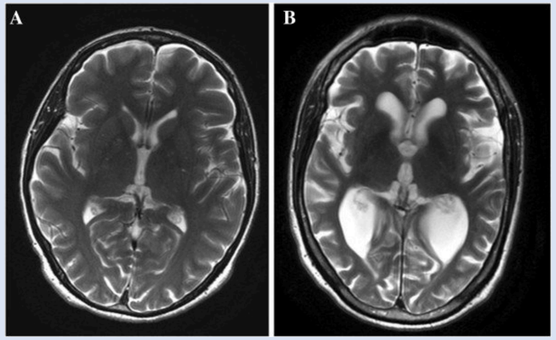
Any technique/method that alllows machines to mimic human behaviour

What is Machine Learning?

Three main **types** of Machine Learning

Supervised

ML approach that uses labeled data to learn and predict outcomes

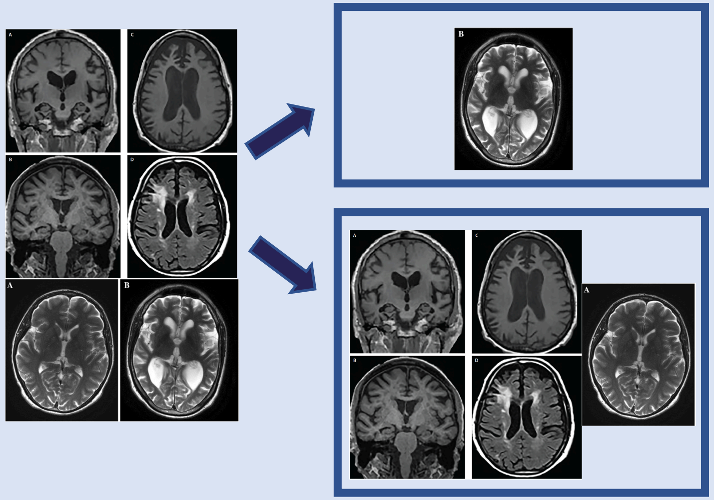


Alzheimer Healthy

Task driven

Unsupervised

ML approach that uses unlabeled data



Data driven

Reinforcement Learning

ML approach

Beyond the scope of this lecture

Learning from mistakes

Cornelius et al, 2013

Ayers et al, 2019

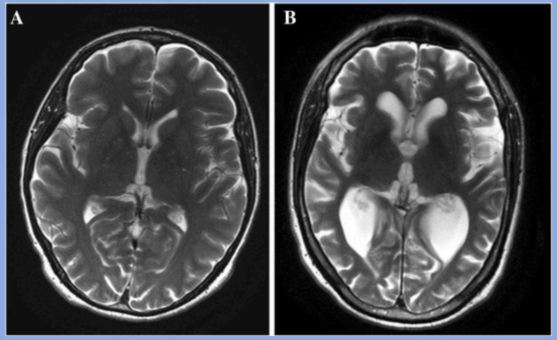
Imperial College London

What is Machine Learning?

Three main **types** of Machine Learning

Supervised

ML approach that uses labeled data to learn and predict outcomes

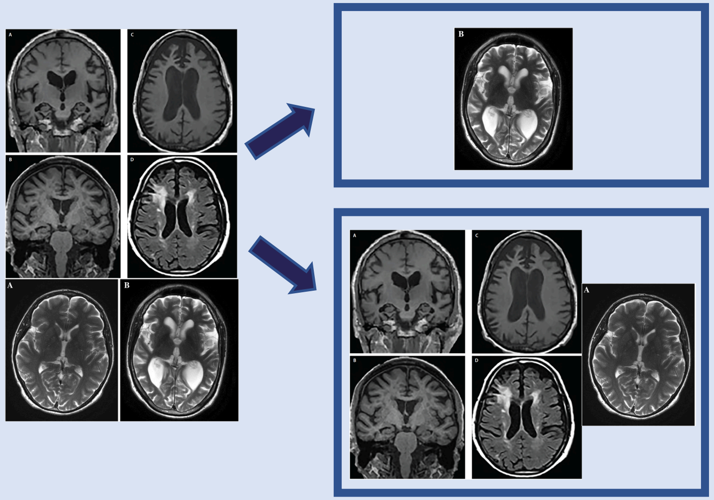


Alzheimer Healthy

Task driven

Unsupervised

ML approach that uses unlabeled data



Data driven

Reinforcement Learning

ML approach

Beyond the scope of this lecture

Learning from mistakes

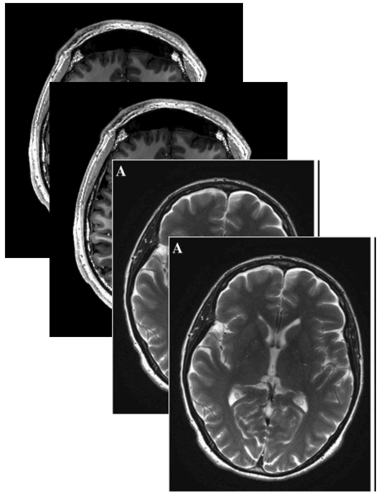
Cornelius et al, 2013

Ayers et al, 2019

Supervised Learning

Supervised learning can be divided into **CLASSIFICATION** and REGRESSION

Input data



Healthy

Healthy

Alzheimer

Alzheimer

...

...

Data

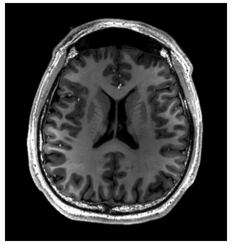
Categorical Label (e.g. disease status)

Model training using ML algorithm

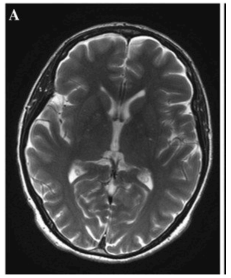
ML model

Output of ML algorithm, can be used to make predictions on unseen data

Prediction



Healthy



Alzheimer

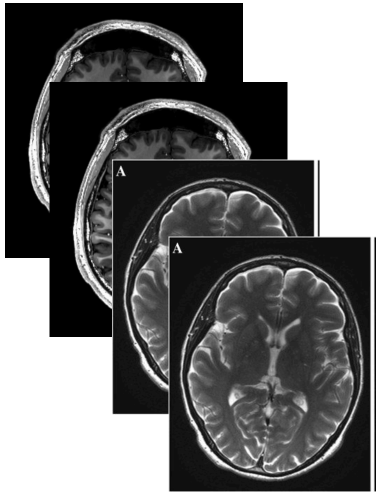
...

CLASSIFICATION: prediction of a categorical outcome label

Cornelius et al, 2013

Supervised learning can be divided into CLASSIFICATION and **REGRESSION**

Input data



70

65

67

72

...

...

Data

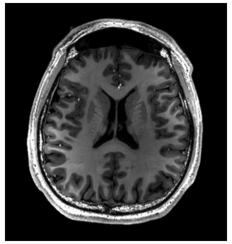
Continous Label (e.g. Age)

Model training using ML algorithm

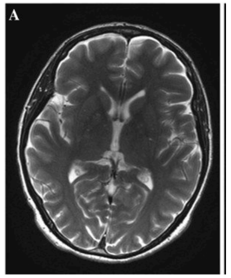
ML model

Output of ML algorithm, can be used to make predictions on unseen data

Prediction



63



74

...

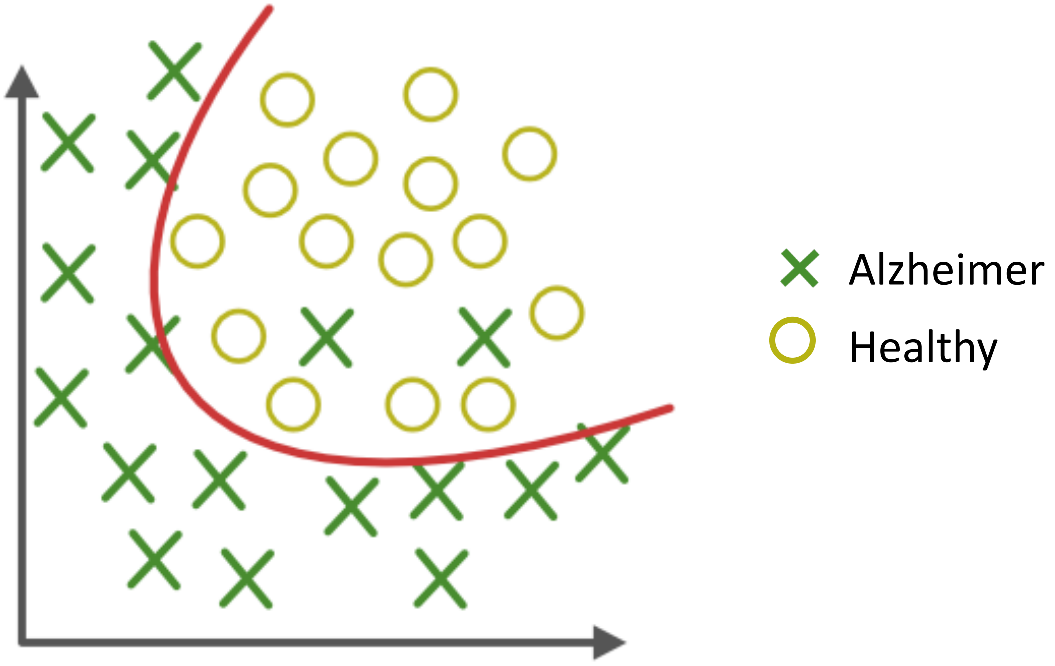
Regression: prediction of a continous outcome label

Cornelius et al, 2013

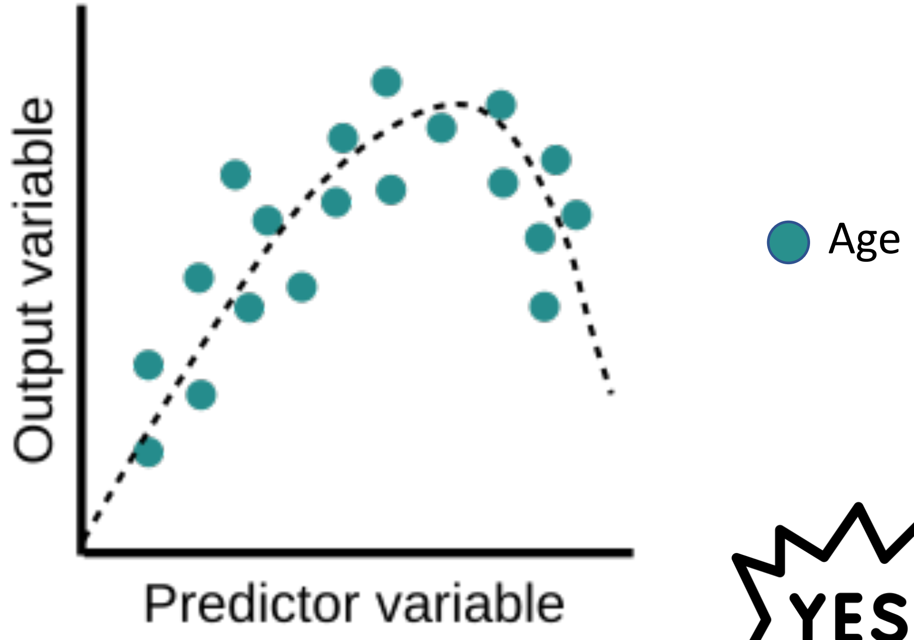
Classification vs Regression

What is the ML model trying to learn? → The ML model tries to learn what is the best line (in red) to separate the different classes (**CLASSIFICATION**) or to fit the data (**REGRESSION**)

Classification



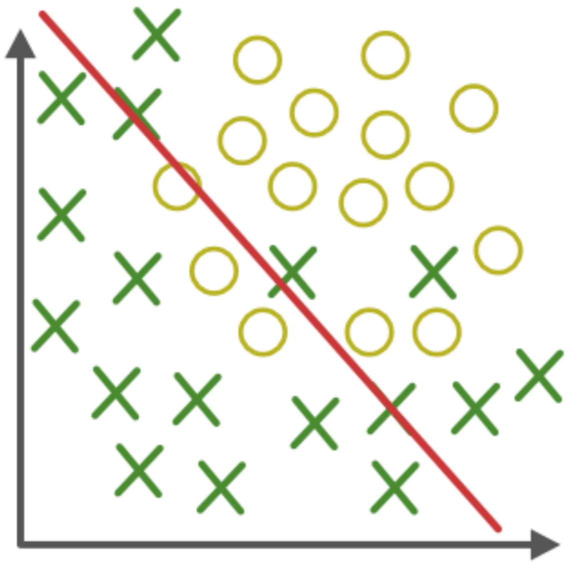
Regression



In real life there is NEVER a perfect split or fit to the data and the best line is hard to find → Can your model be wrong?

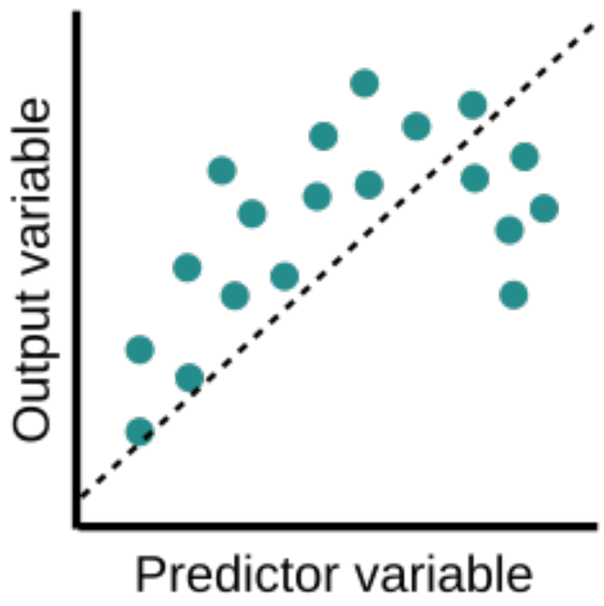
UNDERFITTING: the model cannot properly learn how to predict the categorical or continuos label and performs poorly both on the data on which it was trained and on unseen data

Classification



Alzheimer
Healthy

Regression



Age

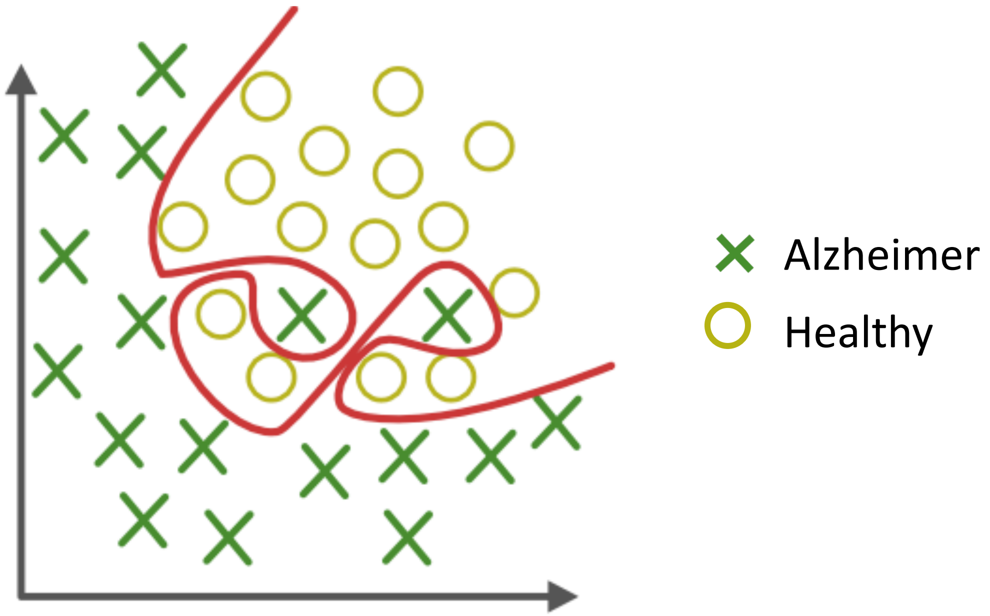
Potential **causes**:
- Not enough training samples
- Data provided are not enough

Images from here

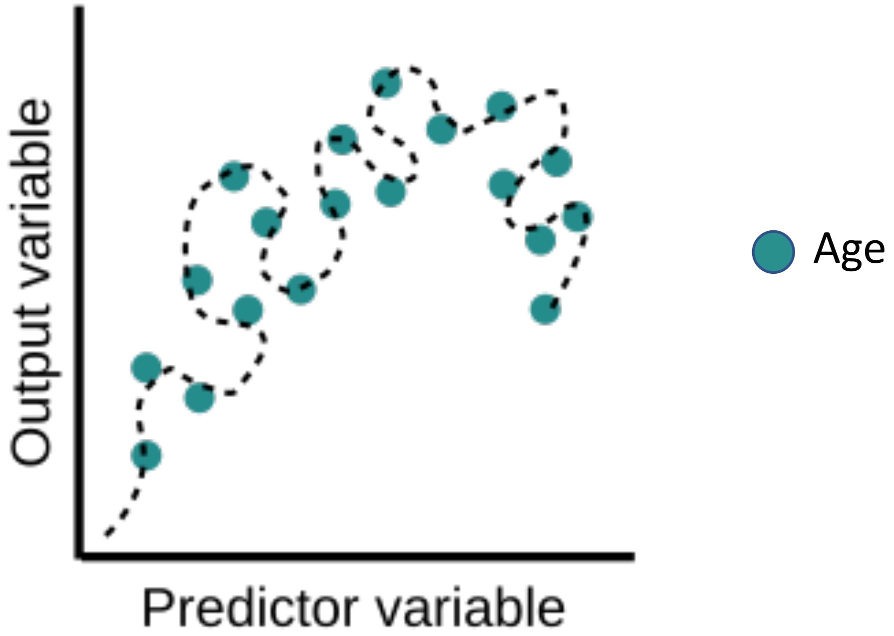
Overfitting

OVERFITTING: the model learns the data on which it was trained *too well* and performs poorly on unseen data

Classification



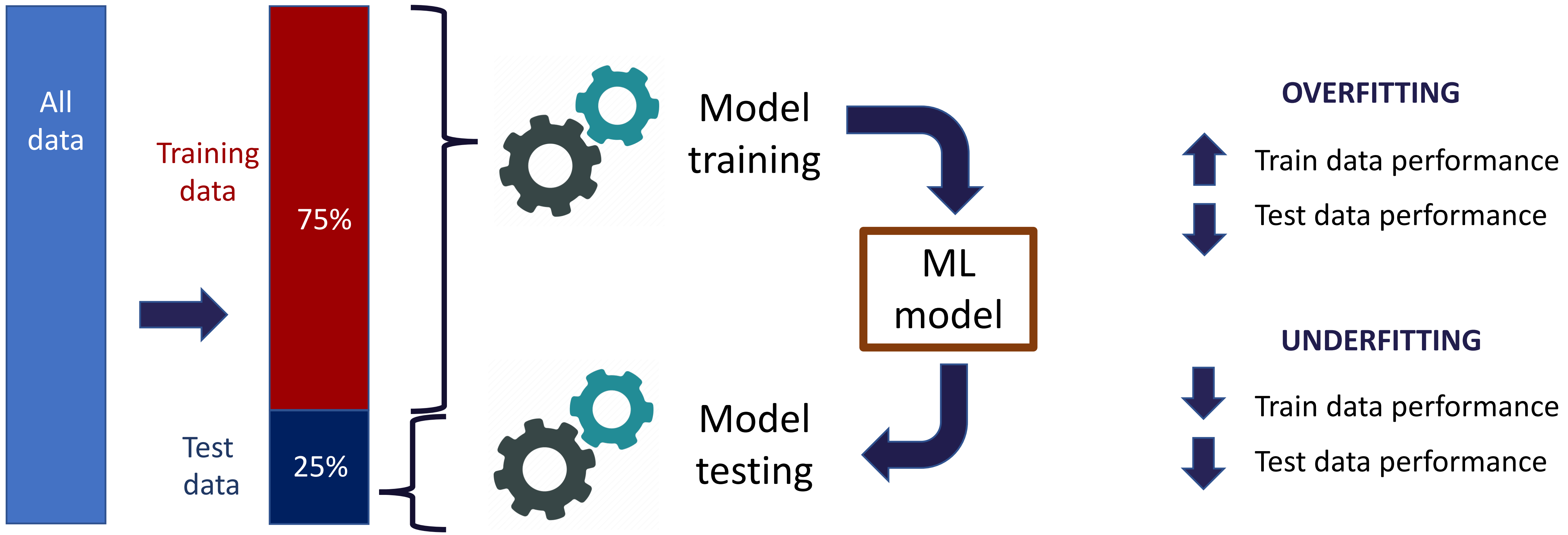
Regression



Potential **causes**:
- Wrong type of model
- Too many features in the input

Is your model overfitted or underfitted?

The standard approach to test whether the model is underfitted/overfitted is to train it on a part of the data and test it on an hold-out dataset



Percentage of splitting depends on size of dataset + ML algorithm used

Predictions on unseen data

K-Fold cross validation

! Splitting in train/test is RANDOM, but based on the splitting performance can change → Some samples are **EASIER/HARDER** to predict than others

K-Fold cross validation: compute the train/test split K times, and calculate the performance on K different test sets

→ Final performance = Average

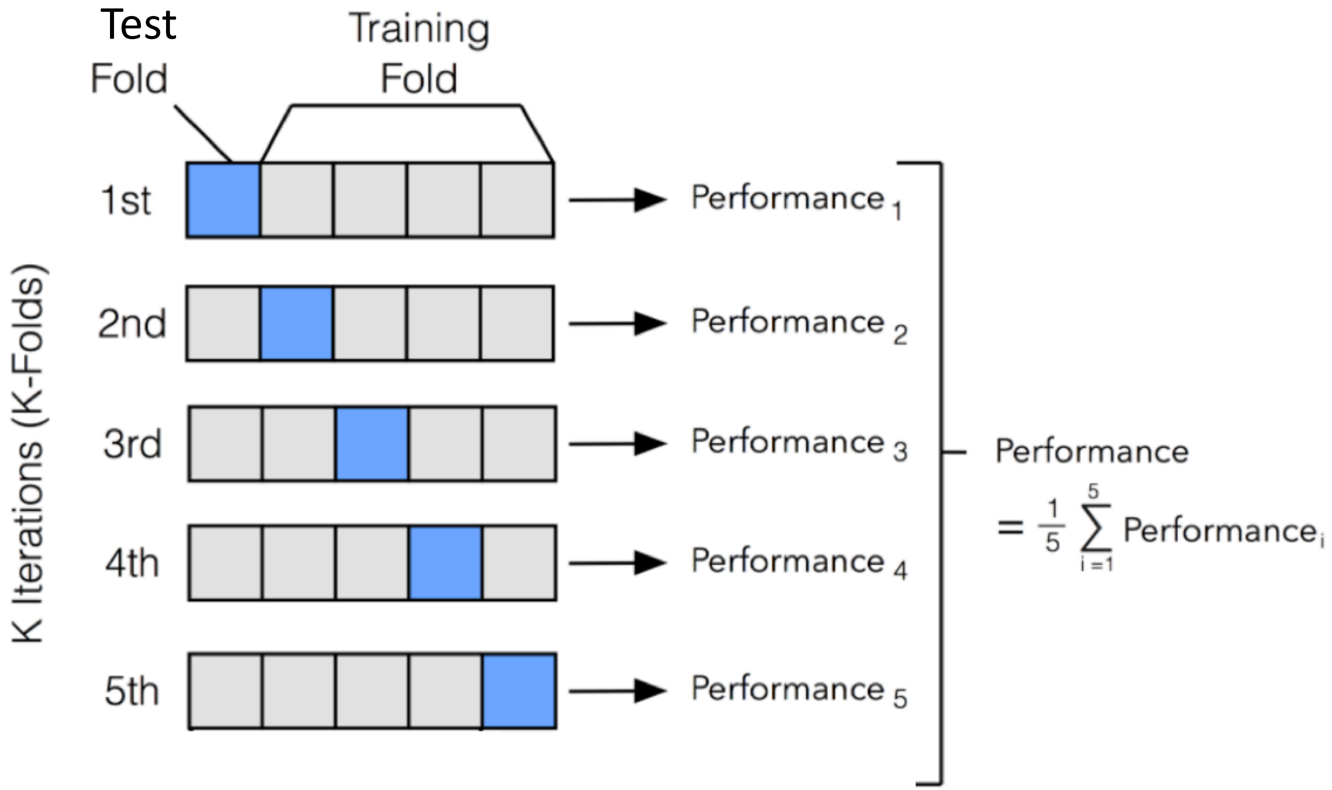


Image from here

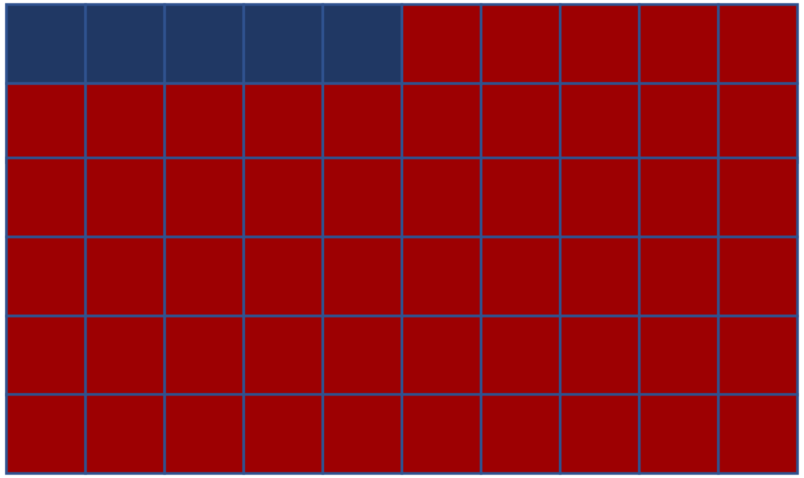
How do you measure performance?

CLASSIFICATION

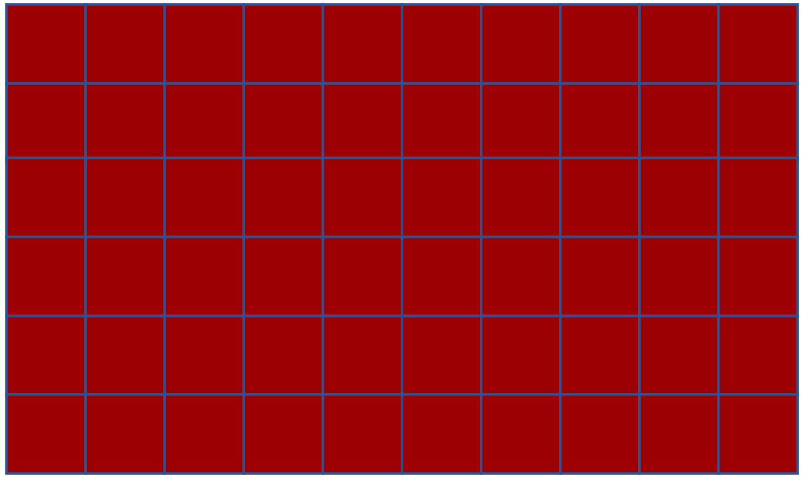
$$Accuracy = \frac{Number\ of\ corrected\ predictions}{Total\ number\ of\ predictions}$$

❌ Not robust to **CLASS IMBALANCE (i.e. unequal number of samples in each class)**

Actual Labels



Predicted Labels



Accuracy:
55/60 = ***0.92***

! **High** accuracy **AND bad** model

Healthy

Alzheimer

CLASSIFICATION

CONFUSION MATRIX → ✓ Robust to **CLASS IMBALANCE**

TP: 0

FP: 0

TN: 55

FN: 5

$$TPR = \frac{TP}{Actual\ Positive} = \frac{TP}{TP + FN}$$

$$FNR = \frac{FN}{Actual\ Positive} = \frac{FN}{TP + FN}$$

$$TNR = \frac{TN}{Actual\ Negative} = \frac{TN}{TN + FP}$$

$$FPR = \frac{FP}{Actual\ Negative} = \frac{FP}{TN + FP}$$

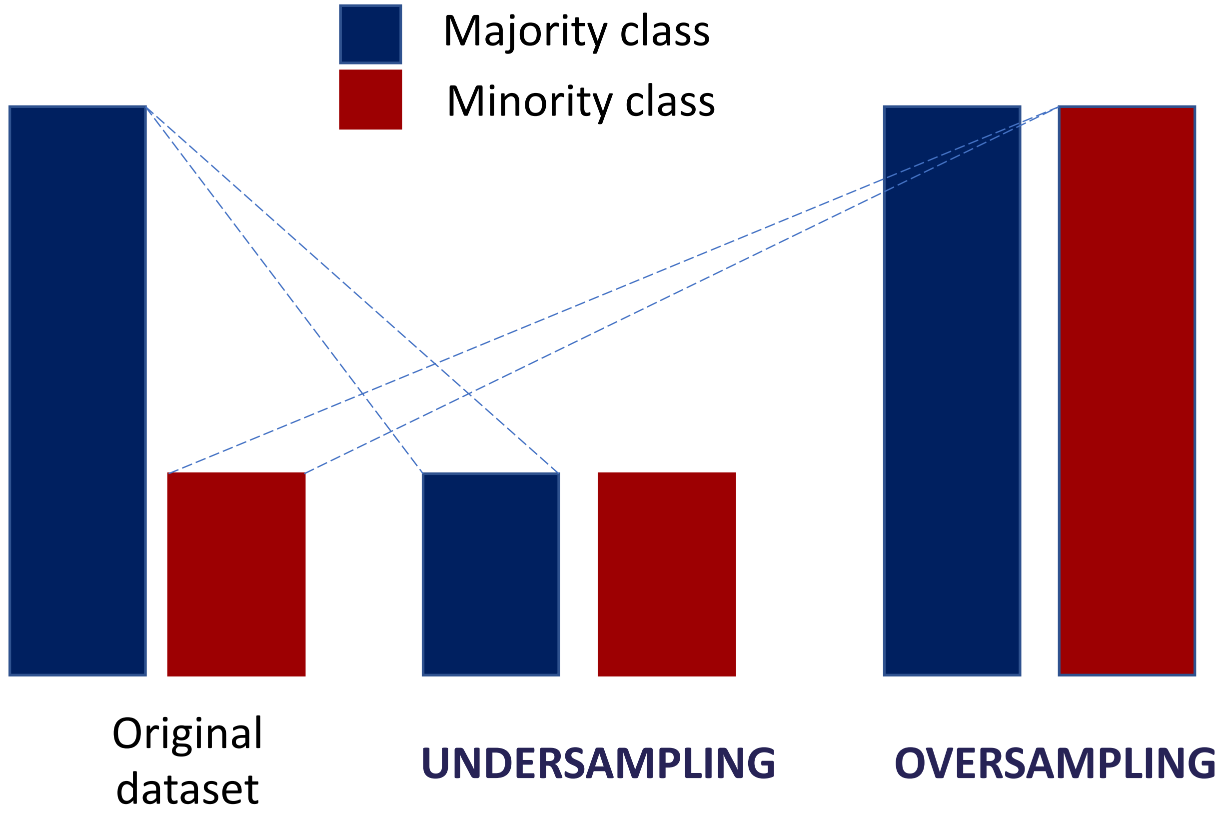
Actual Label

Predicted Label	Alzheimer	Healthy
Alzheimer	True positive	False positive
Healthy	False negative	True negative

Imperial College London

Class imbalance

! **Class imbalance should be addressed DURING or BEFORE model training** → ✗ Model will learn to predict everything as **MAJORITY CLASS** (i.e. class with most samples)



BEFORE

Model **PENALISATION**

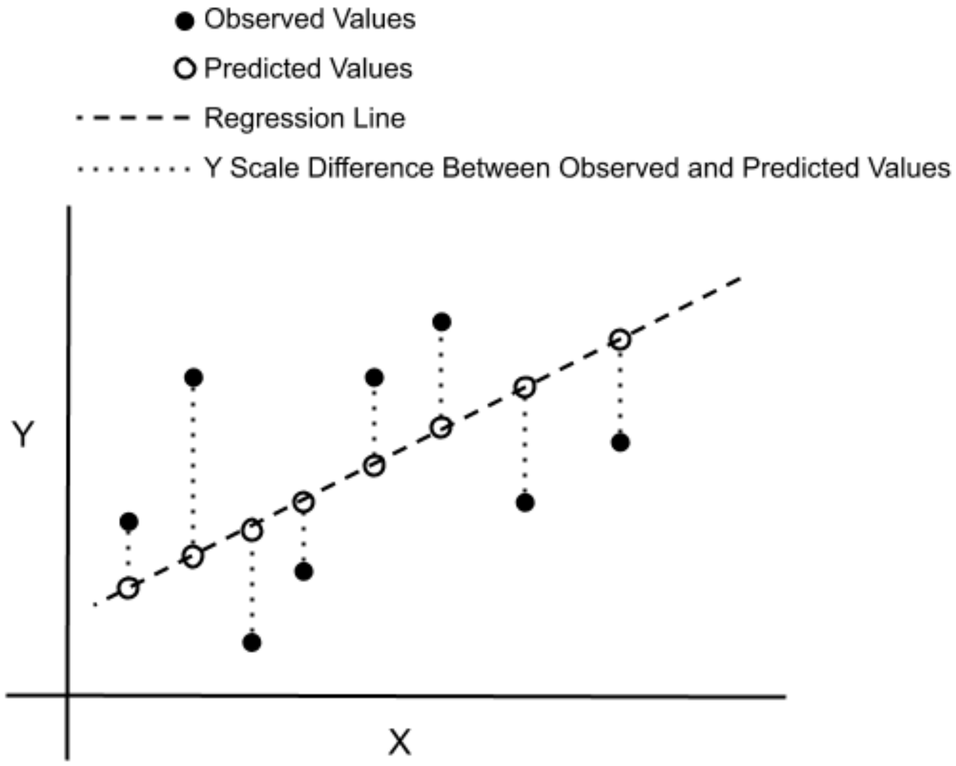
↳ Influence the training of the model so that the model makes fewer mistakes with the minority class

DURING

How do you measure performance?

REGRESSION

! We cannot calculate accuracy in case of regression models



MEAN SQUARED ERROR: measures how close a fitted regression line is to the actual data

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

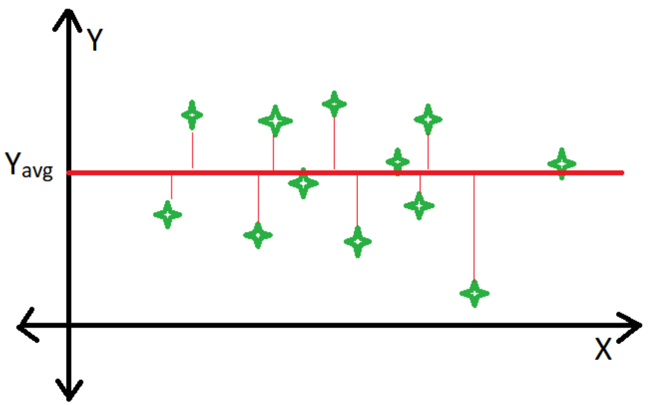
n = number of samples
Y_i = observed values
$\hat{Y}_i$ =predicted values
MSE = mean squared error

➡ Hard to interpret on its own

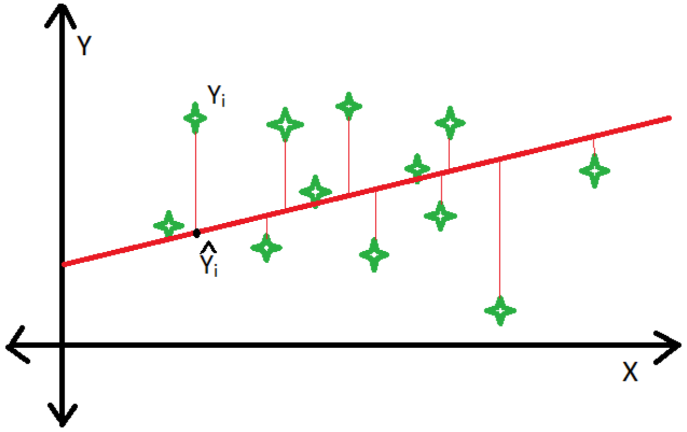
Image from here

REGRESSION

$$SS_{total} = \sum_{i=1}^{n} (Y_i - Y_{mean})^2$$



$$SS_{res} = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$



R2 (or Coefficient of determination): the R2 represents how better your model is compared to the mean model

→ Describes how much of the total variation in your data is explained by the independent variables

$$R2 = 1 - \frac{SS_{res}}{SS_{total}}$$

R2 = R squared
SS_{res} = residual sum of squares
SS_{tot} = total sum of squares

R1 = 1 → Perfect fit

Image from here

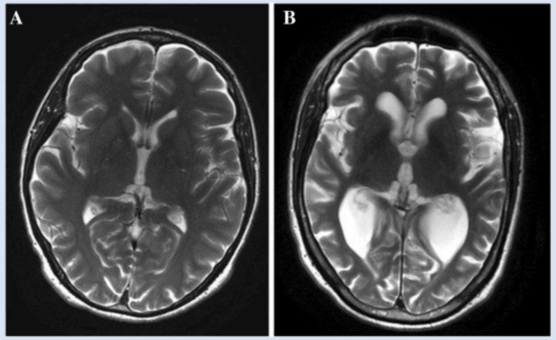
Imperial College London

What is Machine Learning?

Three main **types** of Machine Learning

Supervised

ML approach that uses labeled data to learn and predict outcomes

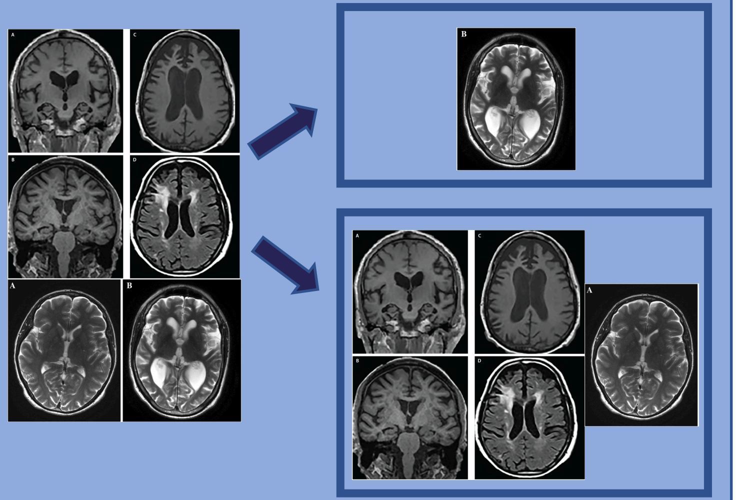


Alzheimer Healthy

Task driven

Unsupervised

ML approach that uses unlabeled data to discover patterns



Data driven

Reinforcement Learning

ML approach

Beyond the scope of this lecture

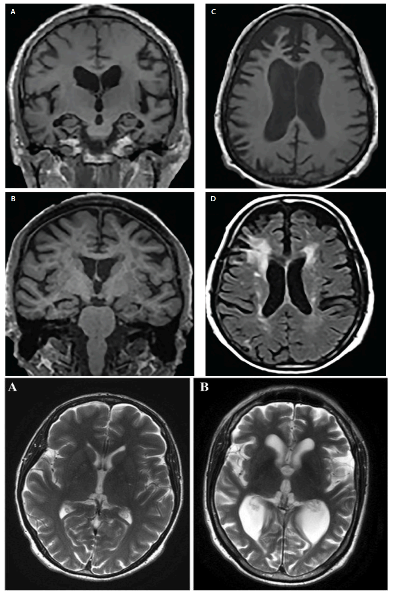
Learning from mistakes

Cornelius et al, 2013

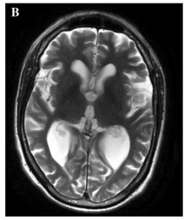
Ayers et al, 2019

Unsupervised learning is commonly used for **CLUSTERING** and DIMENSIONALITY REDUCTION

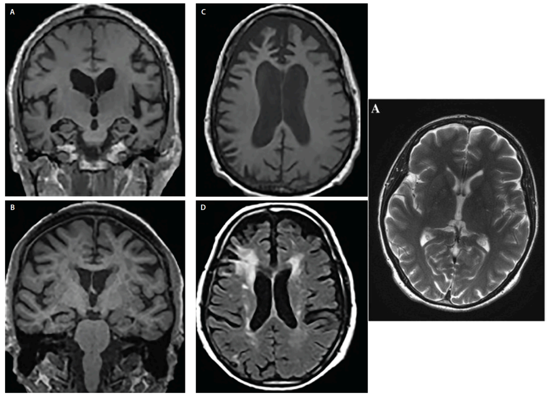
Input data



ML algorithm



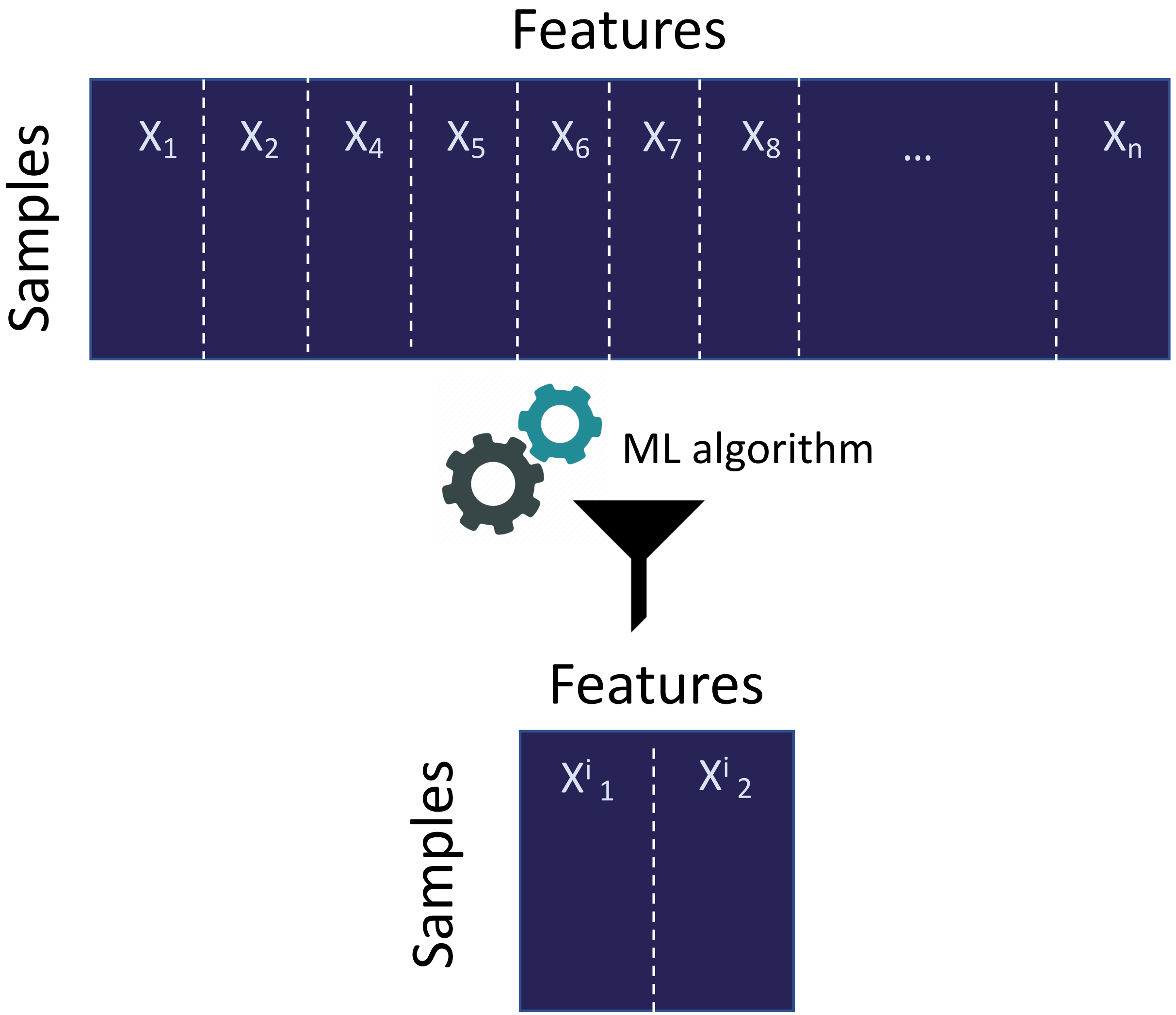
Cluster 1 (Healthy?)



Cluster 2 (Dementia?)

Clustering: grouping of unlabeled data

Unsupervised learning is commonly used for CLUSTERING and **DIMENSIONALITY REDUCTION**



WHY?

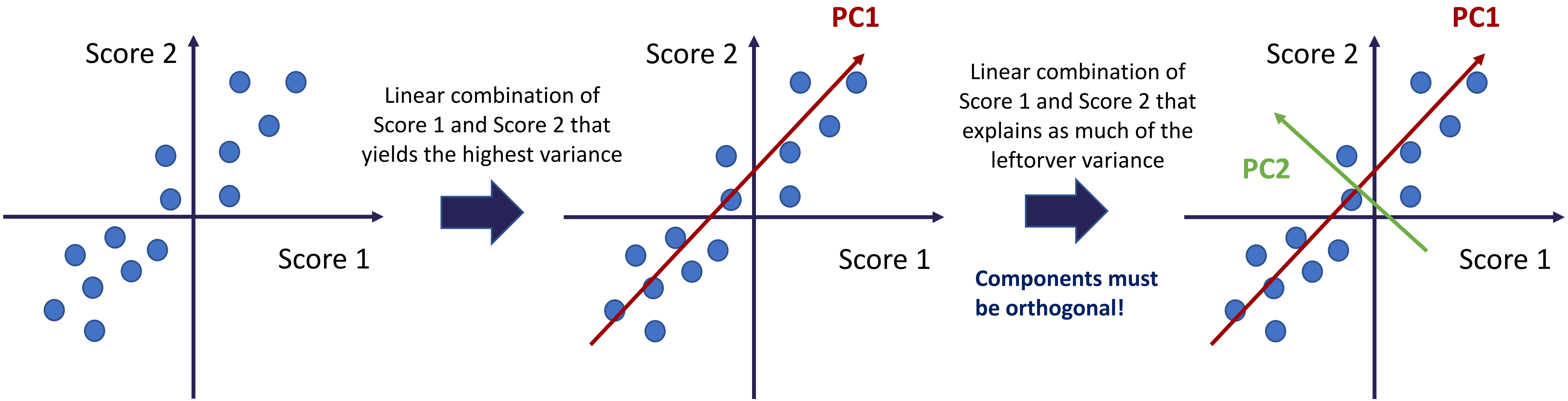
1. Visualisation
2. Decrease model complexity and prevent overfitting
3. Feature selection

! Feature selection differs from DR because it simply selects a subset of features without changing them

Dimensionality Reduction: transformation of data from high to low dimensional space

Principal Component Analysis

Principal Component Analysis (PCA) is a common unsupervised ML technique for dimensionality reduction that transforms a dataset by creating new dimensions that try to explain as much variance in the data as possible.

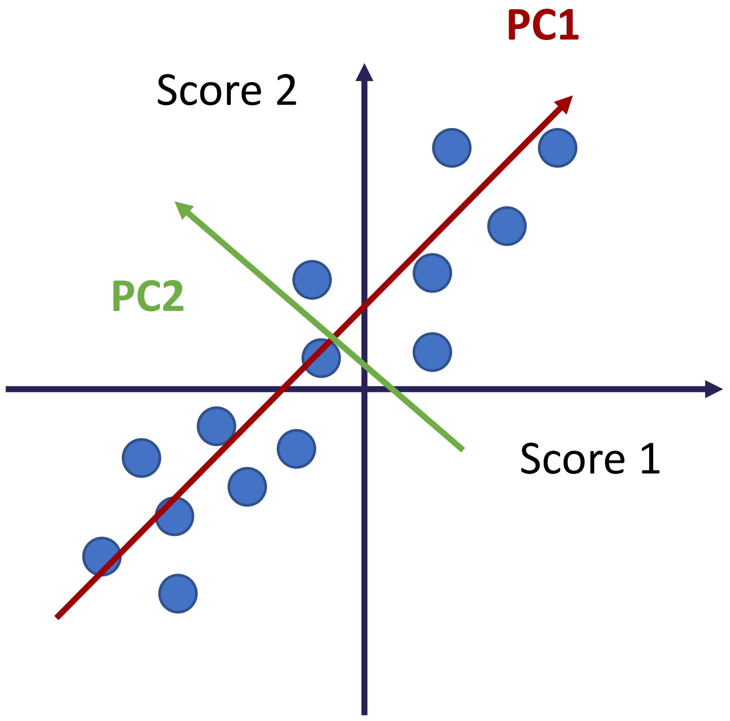


Let's imagine data with only 2 features (e.g. scores in 2 cognitive tests)

PC1 and PC2 are the 2 new lower dimensional features → You can have as many components as features

How do you interpret the principal components?

NOT STRAIGHTFORWARD



Pincipal components are **linear combinations** of the original features

$$PC1 = \boxed{1.5} * Score1 + \boxed{2} * Score2$$

Loadings

Features with higher loadings contributed the most to define the PC

Can be potentially used for **FEATURE SELECTION**

Imperial College London

Questions?