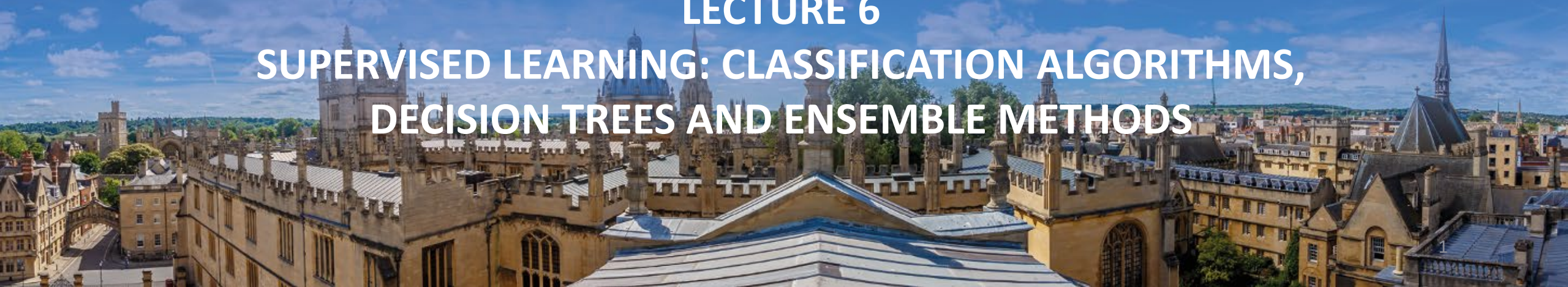
Weekly Oxford Worldwide

DEPARTMENT FOR CONTINUING EDUCATION

UNIVERSITY OF OXFORD

PYTHON PROGRAMMING FOR DATA SCIENCE – PART 2

MASSIMILIANO IZZO & NICHOLAS DAY

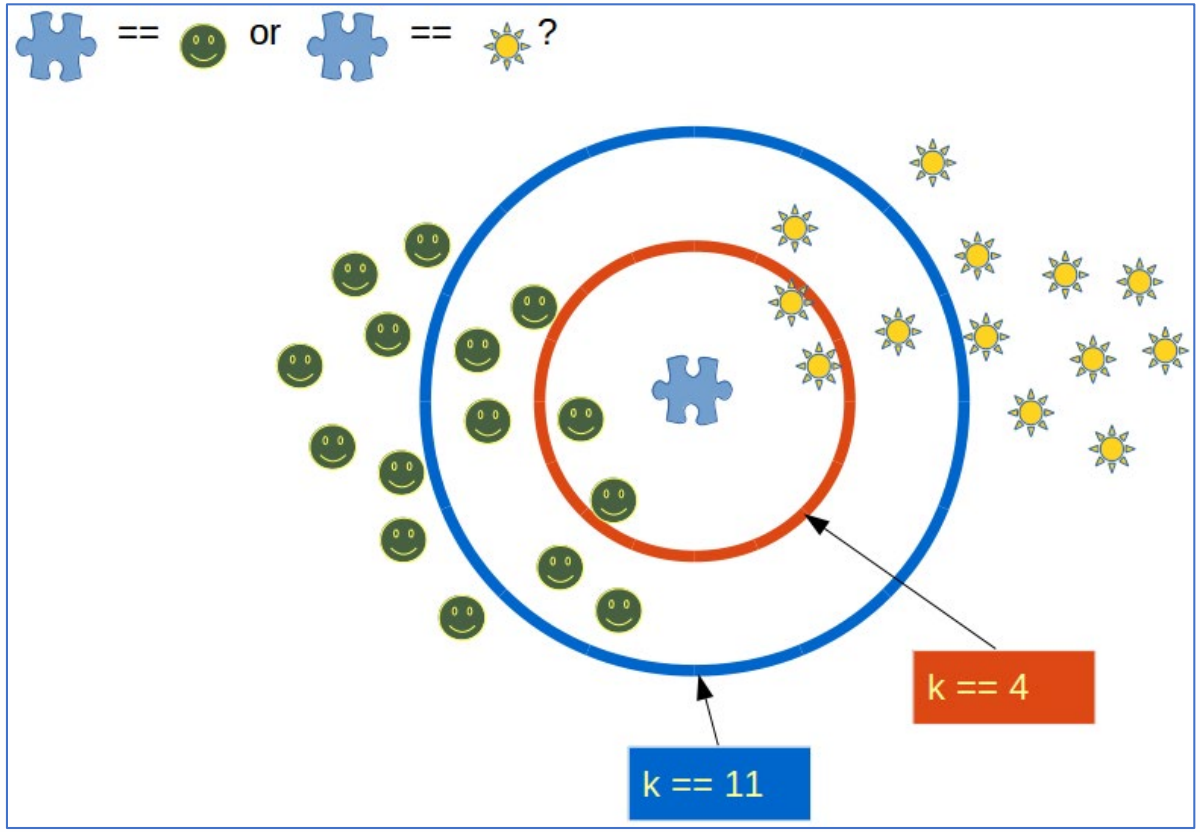


LECTURE 6

SUPERVISED LEARNING: CLASSIFICATION ALGORITHMS, DECISION TREES AND ENSEMBLE METHODS

Nearest Neighbours Classification

- a type of *instance-based learning* or *non-generalizing learning*
- Classification is computed from a simple majority vote of the nearest neighbours of each point
- KNeighborsClassifier
- RadiusNeighborsClassifier
- The closest instances are determined using some distance measure:
  - Euclidean distance
  - Manhattan distance
  - Minkowski distance



source: https://python-course.eu/machine-learning/k-nearest-neighbor-classifier-in-python.php

Naive Bayes

- Naive Bayes methods are a set of supervised learning algorithms based on applying Bayes' theorem with the "naive" assumption of conditional independence between every pair of feature records given the value of the class variable.

$$P(y|x_1,\dots,x_n) = \frac{P(y)P(x_1,\dots,x_n|y)}{P(x_1,\dots,x_n)} \approx \frac{P(y)\prod_{i=1}^{m} P(x_i|y)}{P(x_1,\dots,x_n)} \propto P(y)\prod_{i=1}^{m} P(x_i|y)$$

- So it follows that the most likely prediction for y is as follow:

$$\hat{y} = arg\max_{y} P(y)\prod_{i=1}^{N} P(x_i|y)$$

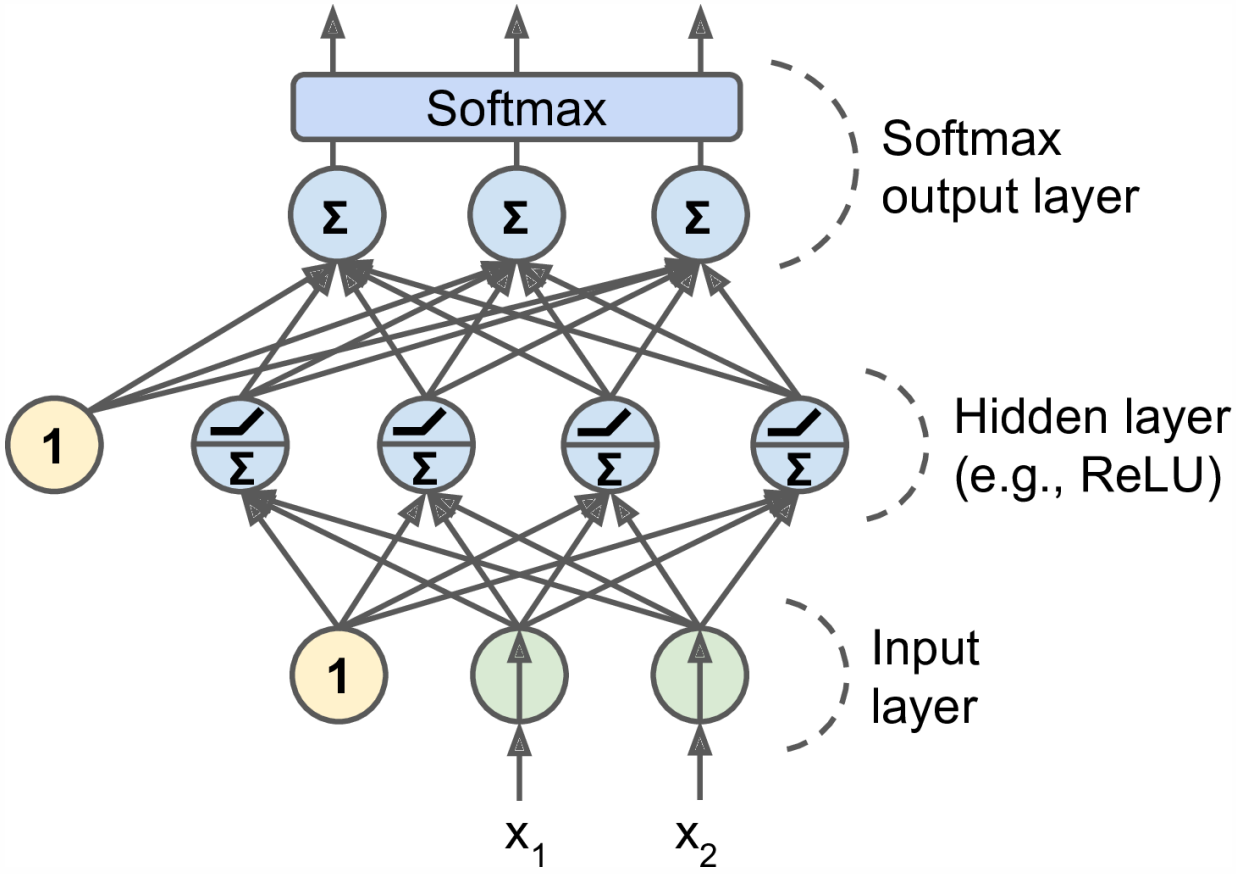
- The different naive Bayes classifiers differ mainly by the assumptions they make regarding the distribution of $P(x_i|y)$: Gaussian, Multinomial, Bernoulli…

Decision Trees

- Decision Trees (DTs) are a non-parametric supervised learning method used for classification (and regression). The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features.
- Advantages:
  - Simple to understand and to interpret. Trees can be visualised.
  - Requires little data preparation (no normalization)
  - Able to handle both numerical and categorical data.
- Disadvantages:
  - Overfitting
  - Learning the optimal DT is an NP-complete problem => heuristics

Neural Networks: Multi-layer Perceptron

- To be seen…



Ensemble methods

- The goal of **ensemble methods** is to combine the predictions of several base estimators built with a given learning algorithm in order to improve generalizability / robustness over a single estimator
  - Averaging methods: Random Forests
  - Boosting methods: Ada Boost, Gradient Tree Boosting
  - Voting Classifier

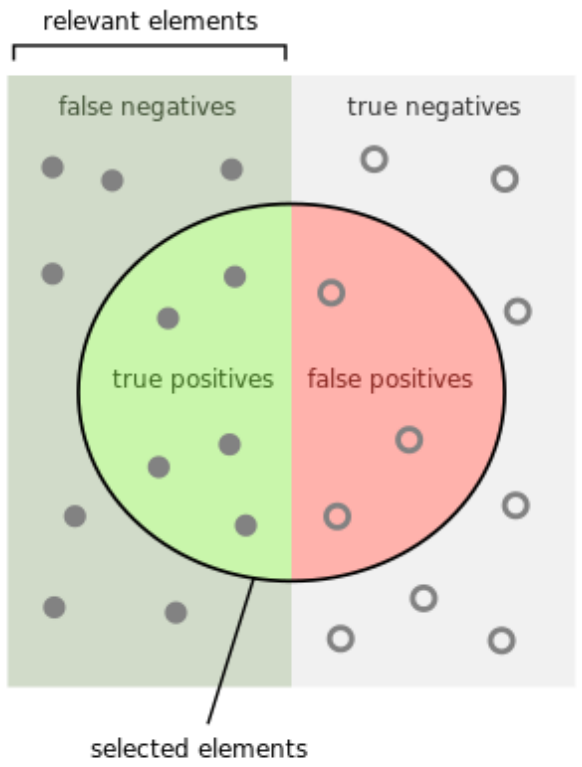
Performance Metrics: Confusion Matrix

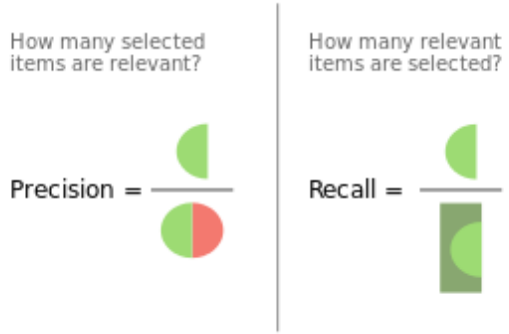
$$accuracy = \frac{correct\ predictions}{total\ predictions} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$F1 = 2\,\frac{precision \times recall}{precision + recall}$$

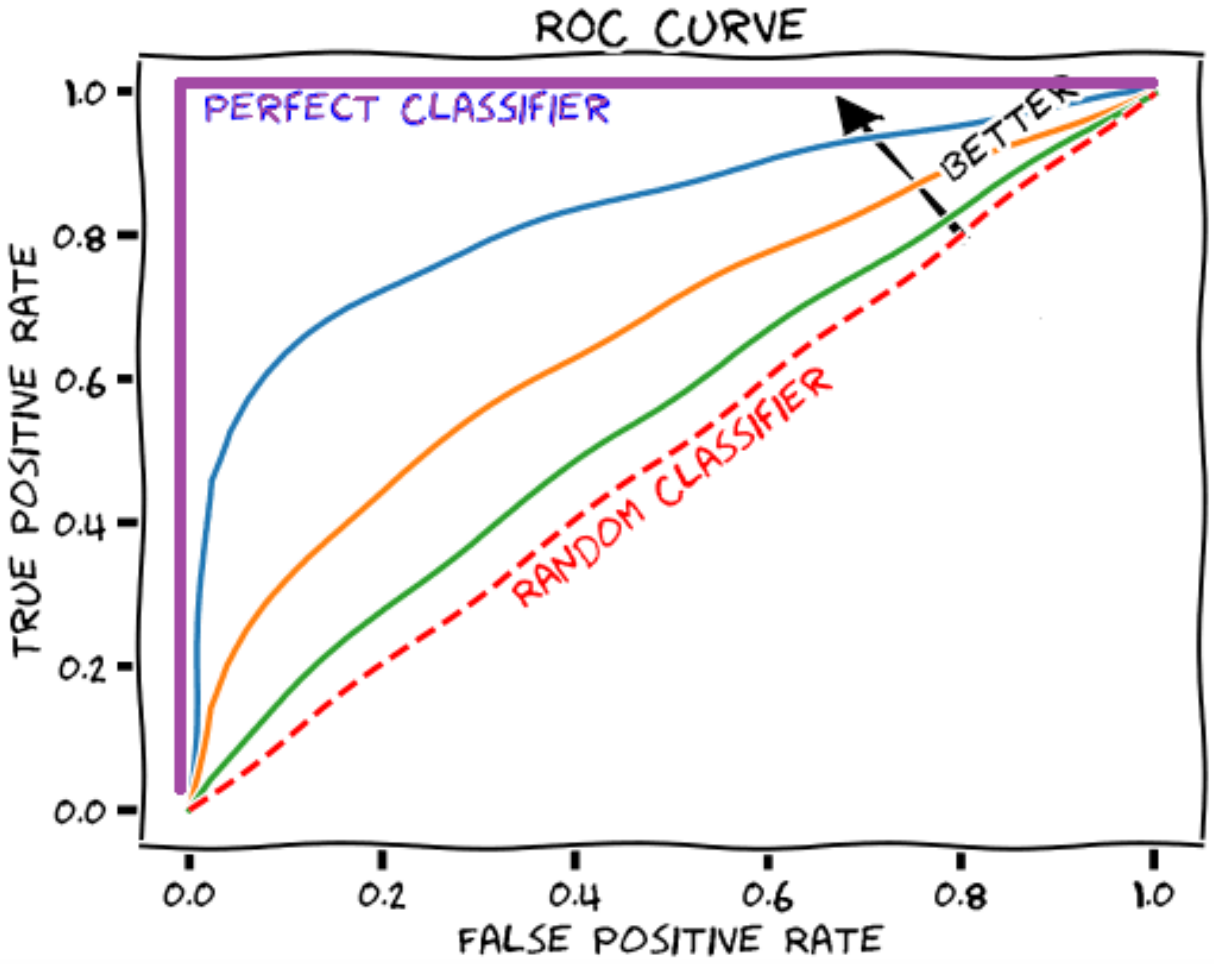




Performance Metrics: Area under the ROC Curve

$$TPR = recall = \frac{TP}{TP + FN}$$

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



https://medium.datadriveninvestor.com/understanding-auc-roc-clearly-explained-74c53d292a02