04 - Logistic Regression

- Logistic regression solves classification problems
- Usually a method of binary classification
- Outcome of dependent variable is discrete
- Assigns probabilities to given outcomes
- Logistic regression uses sigmoid-function
- probability from sigmoid-function
- logit transformation
- Types of regression
  - simple logistic regression
  - Multinomial logistic regression

05 - K-Nearest Neighbour Classifier

- classify examples by assigning them the class of the most similar labeled examples
- very simple but extremely powerful
- Well suited for classifying tasks where the relationships between features are very complex and hard to understand
- training dataset → classified into several categories
- kNN identifies k elemts in the training dataset that are the "nearest" in similarity
- unlabelled test example is assigned to the nearest k cluster

06 - Naive Bayes Classifier

- naive means there is a strong independence assumptions between the given features

- relies heavily on condition probability
- we can decompose the conditional probability on bayes theorem
- choose class with highest probability
- Great for text classification

07 - Support Vector Machine

- Defines margin / boundary → between the data points in multidimensional space
- Goad: find a flat boundary ('hyperplane') that leads to a homogeneous partition of the data
- A good separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class since in general the large the margin the lower the generalisation error of the classifier
- so we have to maximise the margin
- application - classifications or numerical predictions
- pattern recognition - disease classification, text classification and detecting rare events
- Linear separable problem
- Support vectors - the points from each class that are closest to the maximum margin hyperplane // each class have at least 1 support vector
- Convex Hull
- Vector geometry

08 - DecisionTrees

- A type of supervised learning approach
- mainly for classification but can be used for regression
- works fine for both categorical variables and continuous input as well
- split the data/population into two or more homogeneous sets based on significant splitter in input variables
- Root node, decision node, leaf nodes

- How to decide on decision trees
  - Gini index Approach
  - Calculating the information entropy
  - Algorithms based on variance reduction

09 - Random Trees Forest

- Decision Tree tends on over fit
  - every split it makes at each node is optimised for the dataset it is fit to
  - this splitting will process will rarely generalize well to other data
  - Unstable classifiers
- Two solutions
  - Pruning
  - Bagging
- Bias-variance tradeoff
  - bais- error from misclassification in the learning algorithms
    - High bias→ the algorithm misses the relevant relationships between features and target outputs
    - Underfitting
    - Erro due to model mismatch
  - variance - error from sensitivity to small changes in the training set
    - High variance → can cause oevrfitting
    - algorithm models the noise
    - variation due to training sample and randomisation
  - we are not able to optimise both bias and variance at the same time
    - low bias → high variance
    - low variance → high bias
  - Model complexity
- Pruning

- grow a large tree and then prune it back to a smaller subtree
- weak link prunning
- reduce variance

10. Boosting

- can be used for classification and regression
- helps to reduce variance and bias
- bagging create multiple copies of the original data. it consists of several decision trees on the copies and combining all the trees to make predictions. We construct these trees independently.
- **boosting** - the decision trees are grown sequentially so each tree is grown using information from previously grown trees
- these trees are not independent from each other
- boosting is a sequential learning algorithm
- a weak learner is not able to make good predictions
- combining weak learners can prove to be an extremely powerful classifier
- by fitting small trees (decision stumps). we slowly improve the final result in cases when it does not perform well
- Next level is adaptive boosting algorithm

11 - Clustering

- Principal component analysis (PCA)
  - PCA gives us al low dimensional representation of a dataset
  - able to find linear combinations of features / variables that are mutually uncorrelated
  - Linearly uncorrelated variables are the principal components
  - good for visualisation
  - Can be done
    - eigenvalue decomposition of a data covariance / correlation matrix

    - singular value decomposition of a data matrix usually after mean centering
- K-means algorithms