# Assignment

## Decision Tree

## Random Forest

Multiple learning models are combined to increase the classification accuracy. The combination is done via bagging, which creates from a set of noisy and unbiased models, a model with low variance. Because of that the random forest algorithm works as a large collection of decorrelated decision trees. From a matrix $S$ containing all features as columns and the classification for all samples, random subsets $\{S_1, S_2, S_3, \dots, S_m\}$ are created. For each of these subsets an individual decision tree is generated. The complexity of the random forest algorithm is mainly based on the algorithm used to generate the decision trees. So many decision trees build a forest. Afterwards each sample is predicted by each decision tree. The class with the highest occurrence is the prediction of the random forest algorithm.

- What happens, when two classes are predicted with the same occurrence?
  - Randomly choose the class is the standard way in statistics.
- Reference: Random Forest by Leo Breiman

## Idea

- Choose t-SNE for visualization
- Create new data via PCA
- Use random forest for prediction