**Classification Lab**

What is a decision tree?

Supervised learning method.

Flowchart-like structure where each node is a test on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label in other words the decision taken after computing all attributes. The paths from root to leaf present classification rules.

**What is supervised and unsupervised learning?**

* Supervised learning: learning with the desired output (Classification, Regression)
* Unsupervised learning: learning without the desired output (Clustering, Dimensionality reduction, Association Rules/Recommender Systems)

Supervised learning is simply a process of learning algorithm from the training dataset. Unsupervised learning is modelling the underlying or hidden structure or distribution in the data in order to learn more about the data.

**What is pruning? What is it used for?**

It is a technique used to reduce the size of decision trees by removing sections of the tree that provide little power to classify instances. Pruning reduces the complexity of the final classifier, and hence improves predictive accuracy by the reduction of overfitting.

Overfitting is the result of an overly complex model with too many parameters which can became inaccurate.

**What is the CP parameter?**

It is used to control de size of the decision tree and to select the optimal size. if the cost of adding another variable to the decision tree from the current node is above the value of cp, then tree building does not continue.

**Accuracy?**

The number of correct predictions made divided by the total number of predictions made.