**Support Vector Machine (SVM) :**

Support Vector Machines are machine learning models that are mainly used to solve classifications problems.

* Working:

SVM’s use Support Vector Classifiers to classify their data. A support vector classifier or a soft-margin classifier, in contrast to a hard-margin classifier, allows misclassification of the data. With a hard-margin classifier, each datapoint must lie at the correct side of the margin.

* + Soft-margin

For SVM’s to be able to deal with non-linearly separable data, we use the hinge loss function:

is the target of the i-th element. It can be either 1 or -1.

The function will be 0 for data that is classified correctly, this means that the data is on the correct side of the margin. When data is on the wrong side of the margin, the value will increase in proportion to the distance from the margin. Therefore, data that is correctly classified will not affect the loss function at all, and data that is strongly misclassified will increase the loss function by a lot.

To find the best possible soft-margin, this loss function has to be minimalized. This can be done by using cross-validation for example.

* + SVM classifier

The SVM is particularly useful when it is hard to find a soft-margin to classify the data well. When this is the case, no matter which soft-margin we choose, the amount of misclassified data will be great. To solve this problem, SVM classifiers use a Kernel function to bring the data into a higher dimension, and it then uses a higher dimension soft-margin to classify the data. Different functions exist that can be used as kernels.

**Random forest:**

Random forest is a supervised learning algorithm used for classification. It is a technique that uses decision trees, however it is more precise than normal decision trees.

* Working

Random forests work by first creating a bootstrap of the dataset. To do this, we randomly select datapoints from the original dataset. One sample can be picked more than once. After this, create a decision tree using a random subset of the variables in each step. This process of creating a random tree is done a certain number of times. This is a parameter that can be chosen, in sklearn it is 100 by default.

After the trees have been made, we classify new datapoints by using all the decision trees and choosing the average prediction value.

**Cross-validation**:

Cross-validation is a method that is used to compare different machine learning algorithms and find the best one.

* Working:

It works by splitting the data into k folds. The number of folds can be chosen, but usually it is set to 10 (= 10-fold cross-validation). Each time, one of the folds will be used as testing set or validation set, and we will repeat this process until each fold has been the testing set or validation set. Let’s say we wanted to test the accuracy of the algorithm. At the end, the average accuracy over each iteration will be returned. It is possible to use cross-validation to test various things.