SUPERVISED MACHINE LEARNING.

Supervised learning is used whenever we want to predict a certain outcome from a given input, and we have examples of input/output pairs. We build a machine learning model from these input/output pairs, which comprise our training set. Our goal is to make accurate predictions for new, never-before-seen data. Supervised learning often requires human effort to build the training set, but afterward automates and often speeds up an otherwise laborious or infeasible task.

CLASSIFICATION AND REGRESSION.

The goal of classification is to predict the class label which is the choice from a predefined lists of possibilities, classification can be either binary classification dealing with exactly two class or multiclass classification example if you want to predict diabetes disease in a person that means there is more than two class, a person having and not having diabetes where having diabetes can father be divided into types of diabetes. The goal of regression is to predict a continuous number or a floating point number.

GENERALIZATION, OVERFITTING AND UNDERFITTING.

In supervised learning we want to build a model on the training data and then be able to make accurate prediction on new, unseen data that has the same characteristics as the training set that we used. If the model is able to make accurate predictions on unseen data we say it is able to generalize from the training set to the test set, we want to build a model that is able to generalize as accurately as possible.

Generalization refer to how well the concepts learned by machine learning model apply to specific examples not seen by the model when it was learning.

Overfitting means building a complex model for amount of information we have, this occur when you fit a model too close to the particularization of the training set and obtain a model that works well on the training set but is not able to generalize to new data. Underfitting means choosing a simple model.

SUPERVISED MACHINE LEARNING ALGORITHMS.

There are many supervised machine learning algorithms such as linear regression for regression problems, random forest for classification and and regression problems, support vector machine for classification problem, knearest neighbors etc. example the *k*-NN algorithm only considers exactly one nearest neighbor, which is the closest training data point to the point we want to make a prediction for. The prediction is then simply the known output for this training point.