Table of Contents:

1) Machine Learning Introduction

2) What are the types of ML models?

3) What is overfitting and underfitting?

Source: [link](https://github.com/dnyanai/machine-learning-course/blob/master/docs/introduction/intro_to_ML.pdf)

Data:

We have various types of data everywhere.

1) Text

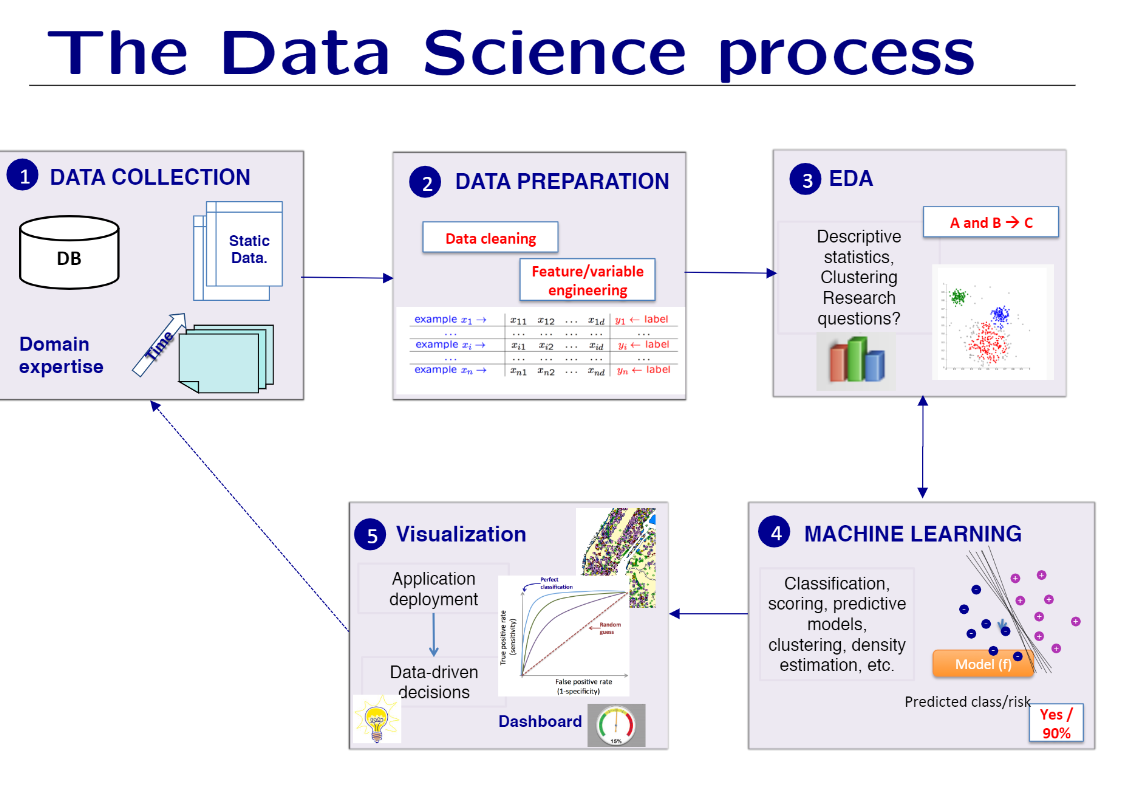
2) Numbers

3) Audio

4) Video

5) Images

Others too like clickstreams...



## Statistical models :

1) Hypothesis testing

2) Experimental design

3) Anova

4) Linear regression

5) Logistic regression

6) GLM

7) PCA

## Supervised Learning:

Supervised Machine learning is a type of machine learning in which the model learns to predict or classify new data based on historical training data or labeled data.

The best example for supervised learning is a teacher and a student learning, supervised model is taught through labeled data about what features to notice to draw a certain type of conclusion. After the model learns on the labeled dataset, it is given a test or never seen data to classify or predict depending on its application.

The features are the properties of the labeled feature which is used to teach the model.

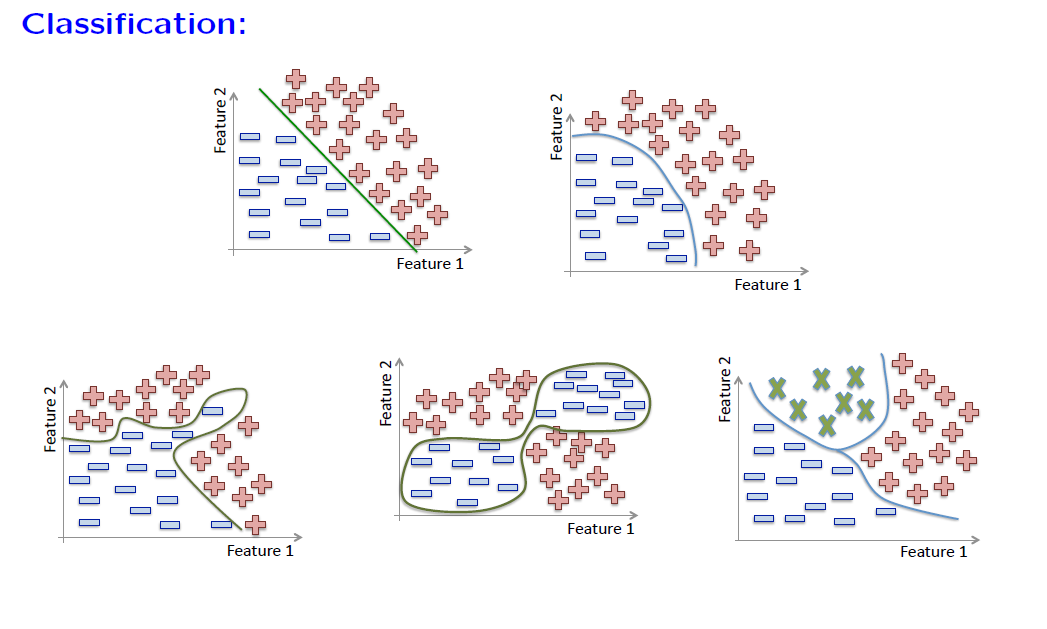
Types:

I) Classification:

Training data – examples/samples with labels.

a. Binary Classifier: y -> discrete i.e., yes/no, 0/1, spam,/no spam

b. Multi Classifier:

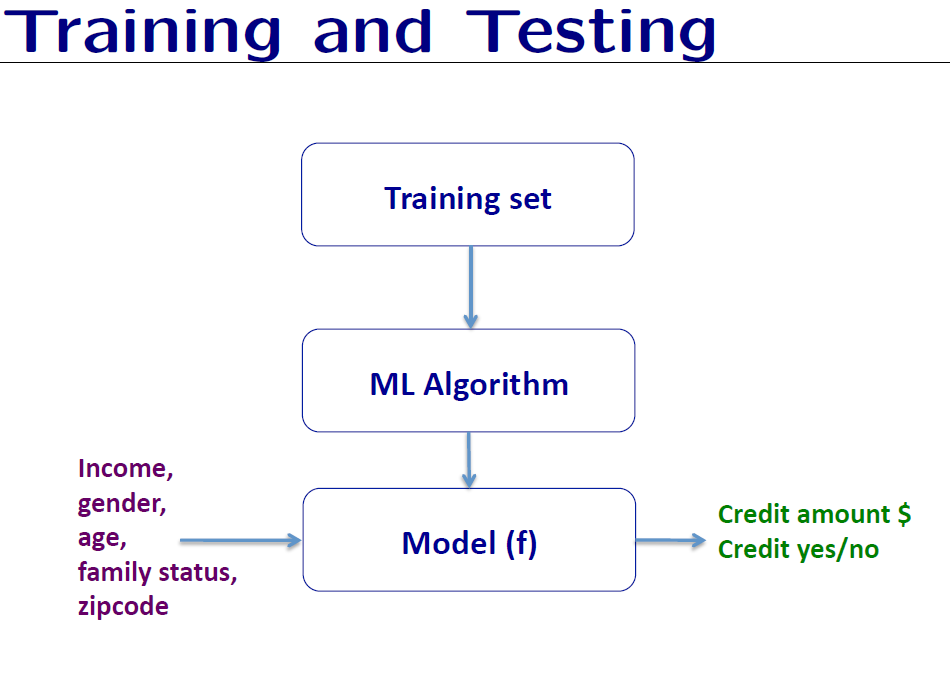


**II) Regression:**

**1) Continuous values or numerical values. Y -> output is a real value,** y element of straight real numbers

**2)** f colon straight real numbers rightwards arrow straight real numbers**, f is called the regressor.**

**3) Credit amount, weight of fruit in function of its length, Income in function of age.**



## ML model I – KNN (K Nearest Neighbors):

1) Does not builds a model, it’s a method. Uses similarity between examples.

2) Assumptions: a. Two similar examples should have the same labels.

b. Assumes all examples are points in the d dimensional space straight real numbers to the power of double-struck d

3) Uses the std Euclidian distance to define the nearest neighbors.

Source: [link](https://github.com/dnyanai/machine-learning-course/blob/master/docs/KNN/Building_a_K_Nearest_Neighbors_KNN.pdf)

Definition: KNN works by taking a data point and looking at the ‘k's closest labeled data points. The data point is then assigned the label of the majority of the ‘k's closest points.

For example: if k=5, and 3 points = green and 2 are red, then the data point to be classified will be green as per the majority.

Colab notebook on a simple example: [Colab\_KNN](https://colab.research.google.com/drive/1zJ821DIKnTlQzs9wQJhHD5xCnvuFoTdM?usp=sharing)

|  |  |  |  |
| --- | --- | --- | --- |
| **What** | **How?** |  |  |
| KNN is an algorithm that uses the similarity between samples. |  |  |  |
|  |  |  |  |

Q1) How is DL different than ML?

- ML needs us to supply the model with a good data representation. This means that the steps of data preprocessing, data transformation and feature extraction and engineering need to be done before we feed data into the model.

- Whereas DL creates a higher-level representation of data as part of the learning process.

Q2 )