Supervised and Unsupervised Learning

Types of Machine Learning:

1. Supervised Learning (Labels Are Defined)

Input and Output data is given to a machine to judge a new incoming input data and predict results on the basis of the older given input and output data that was given to the machine for learning purpose. On the basis of the early data given to the machine the machine will now can understand everything related to that data/problem (etc.) and can give appropriate results to us. This learning that is done by the machine using some input and output data is done using some "learning algorithms".

(Input data + Output data) ---> learning algorithm ---> Model <--- new input data

|

v

Output

List of Supervised Learning Algorithms:

1. Support Vector Machines

2. Discriminant Analysis

3. Naive Bayes

4. Nearest Neighbors

5. Neural Network

Types of Supervised Learning Problems:

a. Regression Problems

It is a type of supervised learning in which we are first given with some input and output data and then we work with non-discrete or continuous (unknown or large or infinite) “values”. We do not work in classes/labels in regression.

Label -> A name of a class.

Example of labels:

Goat, Lion, Deer, Cow, Giraffe, etc. all these are class names(labels)

Each label is represented as a different sign/denotation in the graph but since in unsupervised learning the machine is not given any input and output data so there’s no label defined for the input data that will be given to it to find results of the problem.

Features -> anything using which we differentiate between two or more objects/data is called a feature. In machine learning, we feed the machine with some features (input and output) and then machine predicts/decides/take-decisions based on these features.

Example of features:

In fruits -> Apple color, apple shape, banana color, banana shape etc.

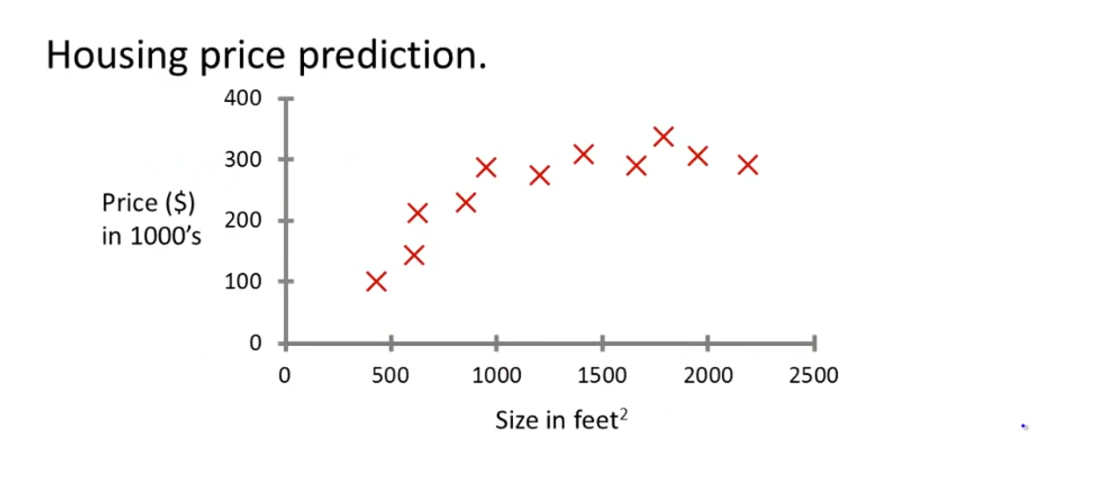
In house and price example -> area in sq feet, price etc.

Examples:

1.You “have a large" inventory of identical items. You want to predict that how many of these items will sell over the next 3 months?

Note: here we “have” (input and output data is already given) large (unknown) or continuous items in our inventory. Therefore, this problem lies under regression problem.

2.



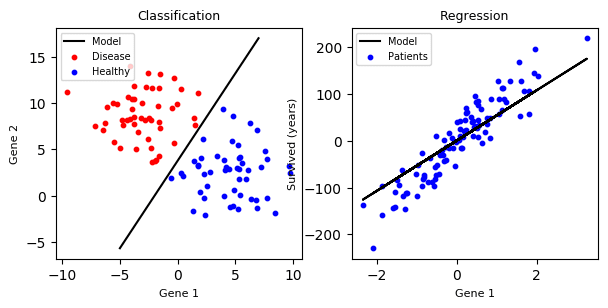
3. Students -> regression -> percentage of marks obtained

Note: here marks percentage is a non-discrete/floating/undermined value of data therefore this also comes under regression.

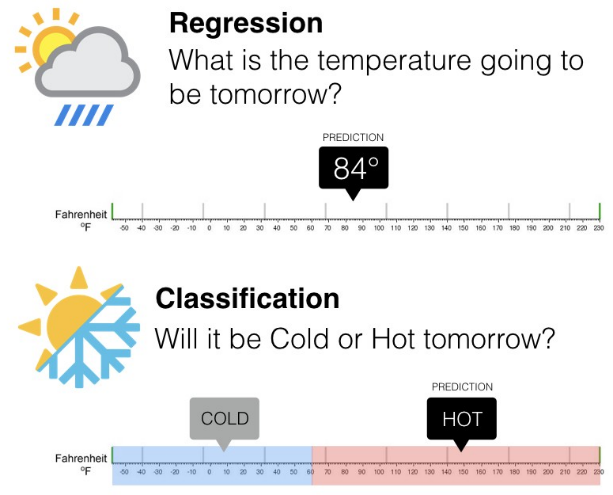
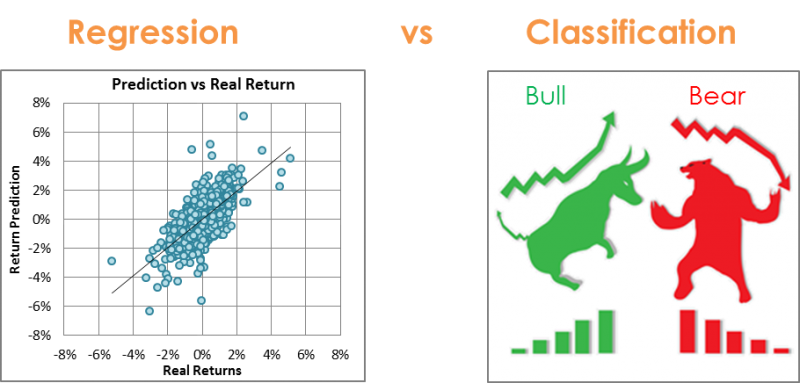
b. Classification Problems (Labels are defined or given)

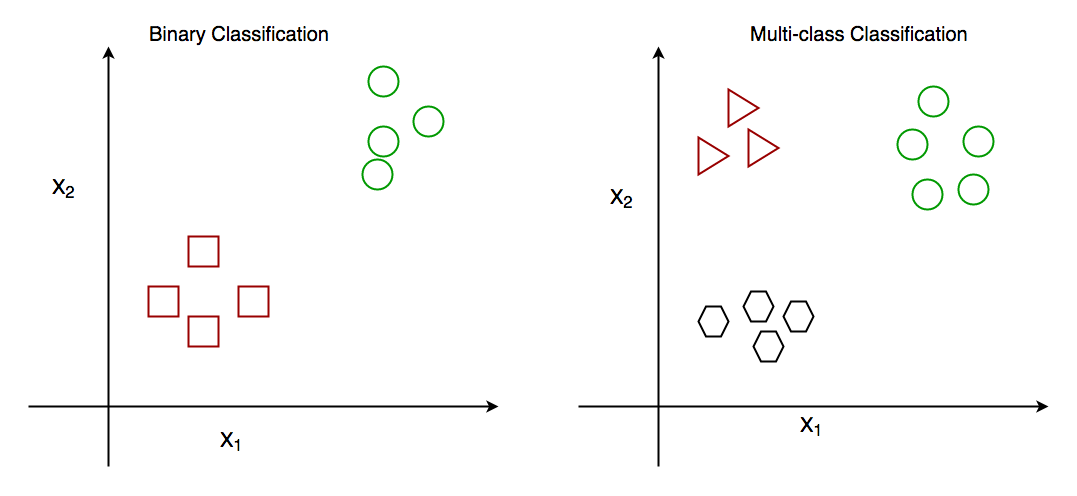
It is a type of supervised learning in which we are first given with some input and output data with labels (denotations/classes) and then we give machine or the learning algorithm some input and then it will identify that in which class/label that given input lies.

Differences Between Regression and Classification:



In the above example, the classification part contains a variety/class/label of data in different color that we have given to our learning algorithm in the first step (feeding input and output data) and now based on this data it will be identifying new input data and will tell us under which class that new input data lies and do predictions based on this given data. However, in regression we also have input and output data but it is not classified/labelled. It has some value, similar or non-labelled (data with no classification) data. Also, if any input data point lies on the division of classes line then the model may give wrong results. No machine learning model is 100% accurate.



2. Unsupervised Learning (Labels Are Not Defined)

“Only” Input data is given to a machine and machine then divides the data into different clusters in such a way that this X data belongs to this Y cluster. we are only given the data set without any other information. we are not told what each data point is, but we are told to find any structure between the data items through a learning algorithm and based on those structures we will classify them into different clusters. The unsupervised learning algorithm will then decide/distinguish that this data X lives in this cluster.

This looks similar to classification but is not same because in unsupervised learning clusters, we first explore the given data because there is no input and output data given to the machine before. So, machine/learning algorithm will first explore the data and divide it into clusters. Also, in unsupervised learning we have labels that are not defined because we are not given any input and output data before the machine or the learning algorithm will be dividing the data into classes after observing/seeing it.

Example of unsupervised learning:

Given a set of news articles found on web, group them into set of articles of same stories.

(this is an unsupervised learning problem because we have no prior knowledge of similarity between the news articles on web, we are directly given the input only and asked to group them (make clusters) and separate them if they have same stories)

Differences between Classification (Supervised) and Clusters (Unsupervised)

