29.1 How Classification works?

Let us consider the Amazon Fine Food reviews dataset. For each review, we have a text associated with it, and this text is converted into a vector form through one of the vectorization techniques like BOW, TF-IDF, Avg Word2Vec, TF-IDF Weighted Average Word2Vec, etc. Now for each review, we have a vector and also the data that tells us whether the review is positive or negative.

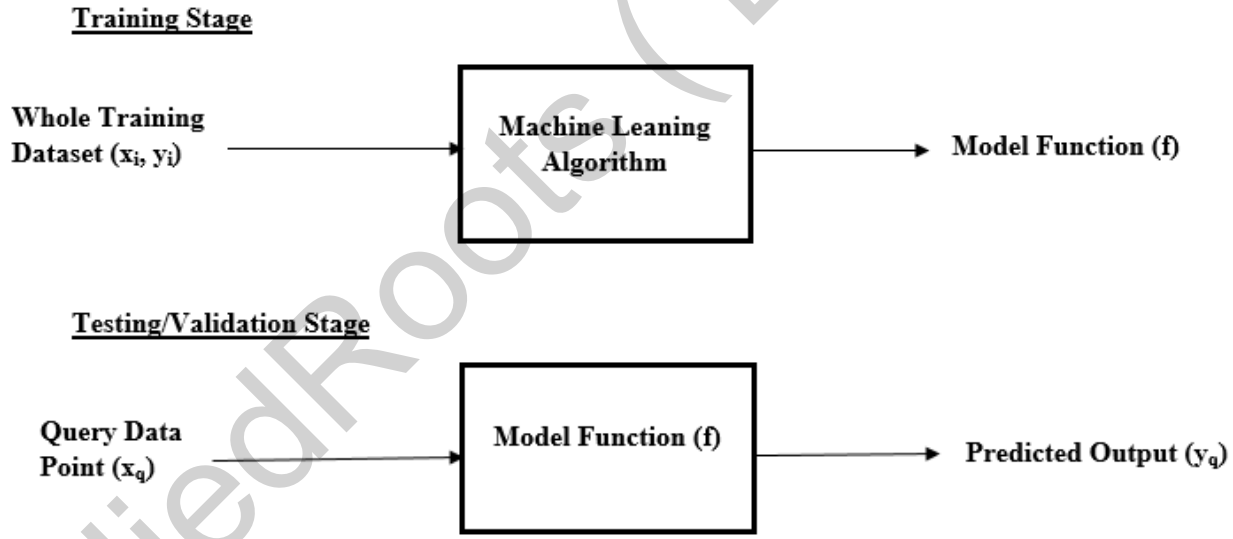
The classification task for this problem is, for a given review, we have to predict if it is positive or negative.

Let us assume each review in our dataset is denoted as 'x_i'. Classification here is given a review 'x_i', we have to find a function 'f' such that if we apply that function on 'x', we get the output 'y' either as 'Positive' or 'Negative'. This is the central concept of classification.

$y_i = f(x_i)$

where $x_i \rightarrow$ query review (in vector form)

$y_i \rightarrow$ predicted class label



The algorithm gets trained using the training data and computes the function 'f'. After computing the function 'f', in the validation stage, this function is used to predict the output for the test data.