

Classification and Regression Models

What is Classification

If we take the example of Amazon food review problem, where we are having C_x reviews. Based on this review is a text. So, at the end we determine that the review is true or false.

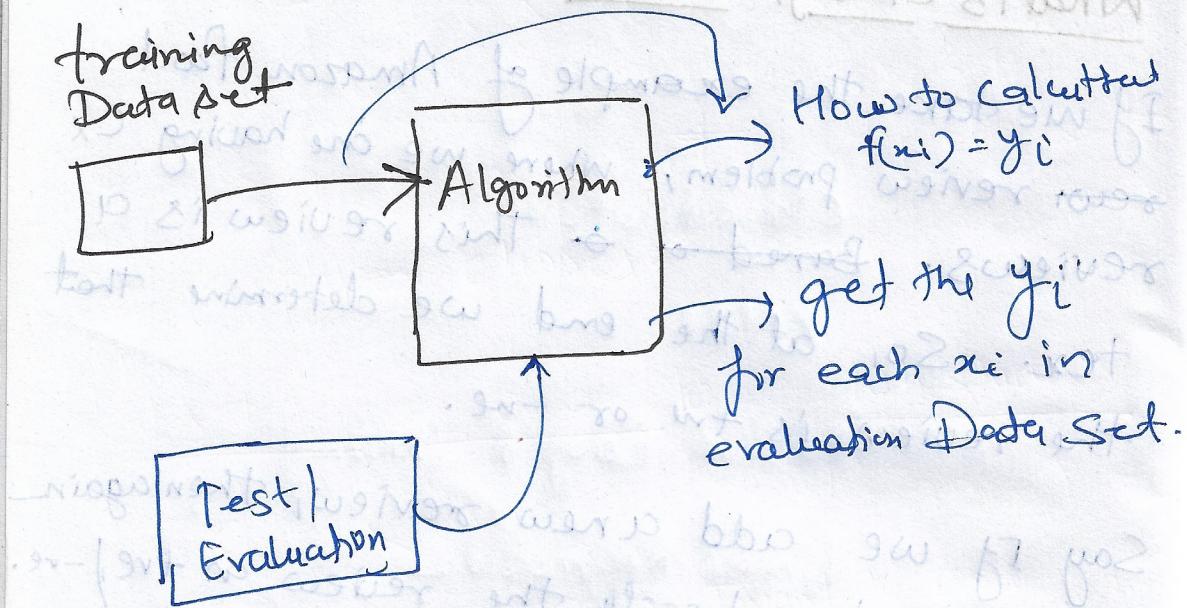
Say if we add a new review, then again we have to classify the review as true.

So, classification can be thought of as finding a function $y = f(x)$

I want to find a function f in which if we put the new input, it should classify that I/P into the ~~desired~~ given class.

This is crux of a classification to find out such magical function.

How classification algorithm works



training Dataset : It contains all the

pair of x_i and y_i . i.e. for each x_i

what would be the y_i

$$f(x_i) \rightarrow y_i$$

The algorithm learns with the help of

classification training data.

This state is known Training state.

$$f(x_i) = y_i$$

Once the training is over. Algorithm learns this, the training phase is over.

Then algorithm goes to test or evaluation stage, where it needs to find y_i for all x_i in the evaluation Data set.

Mathematical Representation of Dataset

Notation of Dataset

My whole Dataset D

$$D = \{ (x_i, y_i) \}_{i=1}^n \mid x_i \in \mathbb{R}^d, y_i \in \{0, 1\}$$

Amazon food review
 dataset
 review review

Classification and Regression

The above dataset is called 2 class classification or Binary Classification.

We can also have multiclass classification problem. Say $y_i = \{0, 1, 2, 3, \dots, 9\}$

10-class classification
 problem

So, we have seen a 2-class and multi-class classification problem.

What if $y_i \in R$ (i.e. the whole ~~Dataset~~ Real numbers)

$y_i \rightarrow$ is no more part of finite set of ~~Problem~~ classes.

Let's take an example of 10K students:

$x_i = \langle \text{weight, age, gender, race} \rangle$

$y_i = \langle \text{height} \rangle$ I want to predict height of a student.

Height is a Real number, no more than 9

class.

So, such problem where $y_i \in R$ is

known Regression Problem.

The only difference b/w classification and Regression is the value of y_i i.e.

$y_i = \{0, 1\} \rightarrow$ classification

If $y_i \in R \rightarrow$ Regression $\xrightarrow{\text{finite class}}$