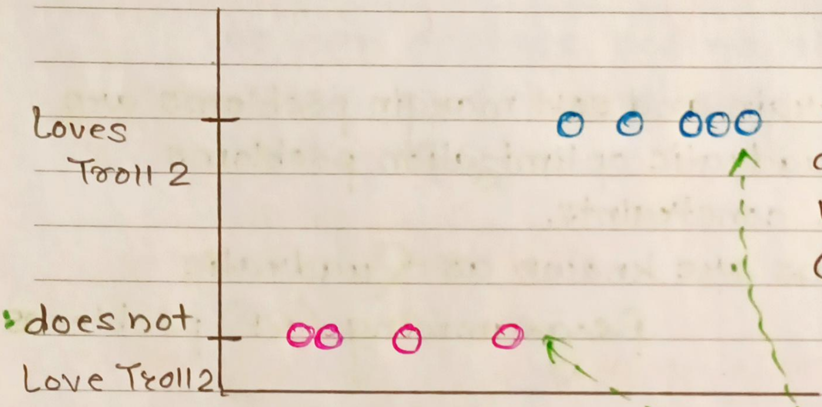
Receiver Operating Characteristic (ROC)

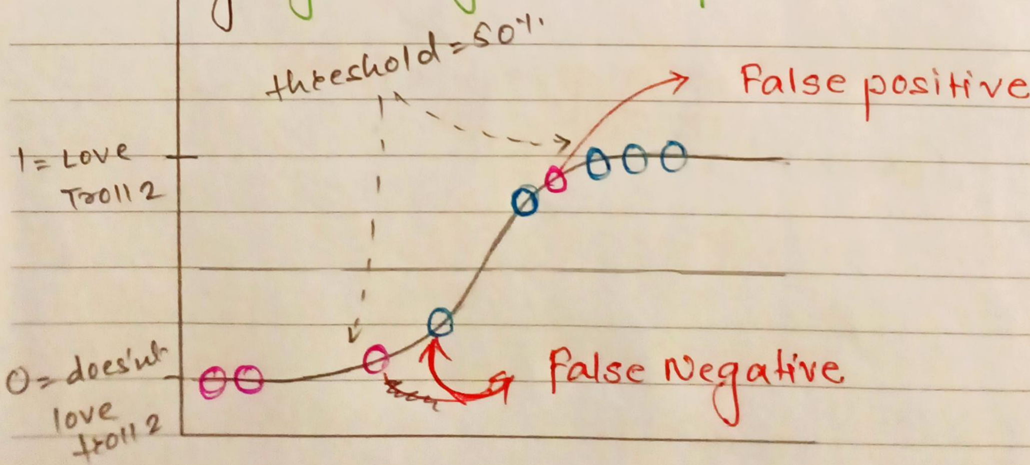
Problem: whether or not someone loves the movie Troll 2?



In this example, we measured the amount of Popcorn a bunch of people ate (grams), which is continuous and whether they Love Troll 2 or Do Not Love Troll 2, discrete.

The goal is to make a classifier that uses the amount of Popcorn someone eats to classify whether or not they love Troll 2.

after using logistic regression to predict →



To find correct identity, we keep on changing threshold value and get different confusion matrix.

Ultimately, we can try any classification threshold from 0 to 1,

and when we do, we end up with 10 different confusion matrix (matrices) that we can choose from.

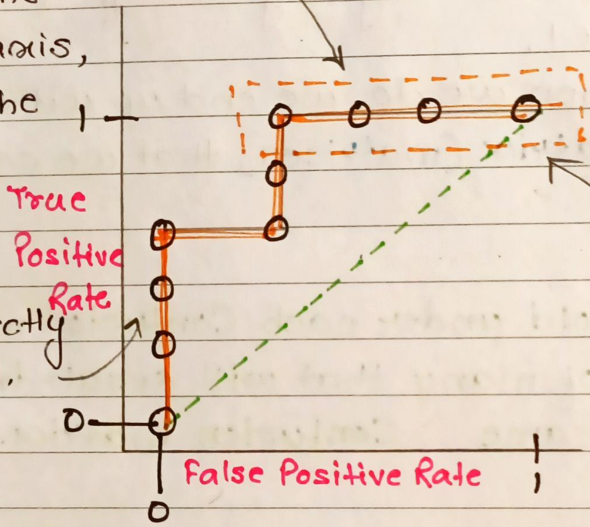
* The threshold under each Confusion matrix is just one of many that will result in the exact same Confusion matrix.

Trying to find the ideal classification threshold among all of these confusion matrices is tedious and annoying.

wouldn't it be awesome if we could consolidate them into one easy to-interpret graph? ⟶ that's when {ROC graphs} came.

- Each black dot on the ROC graph tells us the True Positive Rate and the False Positive Rate for a specific classification threshold.

The higher the dot is along y-axis, the higher the % of actual Positives where correctly classified.



at a glance, we can look at the top row of points and tell that the classification threshold that resulted in the point on the left side performed better than the others because they all have the same True positive rate, but the point on the left has a lower False Positive Rate.

- diagonal line shows, True Positive Rate = False positive Rate.

To construct ROC graph, we will start by using a classification threshold 1. and construct/calculate confusion matrix.

using that confusion matrix, calculate True Positive Rate and False Positive Rate. and then plot that point on ROC graph.

$$\boxed{\text{True Positive Rate (TPR)} = \frac{TP}{TP + FN}}$$

$$\boxed{\text{False Positive Rate (FPR)} = \frac{FP}{FP + TN}}$$

Now, lower the threshold (such as 0.975, 0.965, ...) and calculate confusion matrix for that particular threshold.

Thus, calculate TPR, FPR and plot the points on graph.

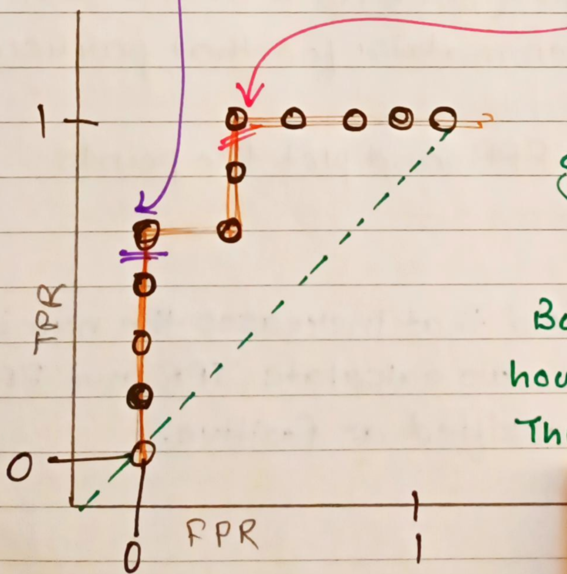
Likewise, for each threshold that increases the number of Positive classification; we calculate TPR and FPR until everyone is classified as Positive.

after we finish plotting the points from each possible Confusion matrix, and connect the dots.

Now, without having to sort through a huge pile of Confusion Matrices, we can use the Roc graph to pick a classification threshold.

If we want to avoid all False Positives, but want to maximize the number of actual Positives correctly classified, we would pick this threshold.

... but if we can tolerate a few False Positives, we would pick this threshold because it correctly classifies all of the actual Positives.



Roc graphs are ~~equal~~ great for selecting an optimal classification threshold for a model.

But what if we want to compare how one model perform Vs another? This ~~is~~ is where

AUC,
Area Under the Curve

comes into picture.

Area Under the Curve

~~AOC~~ AUC → Area Under the Curve.

AUC, measures the entire two-dimensional area underneath the entire ROC curve from (0,0) to (1,1).

The more the area when compared to different model, better the model.