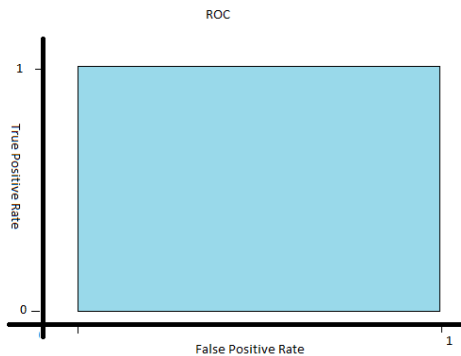


Question 1

2. Explain which model performs better on this data and why?

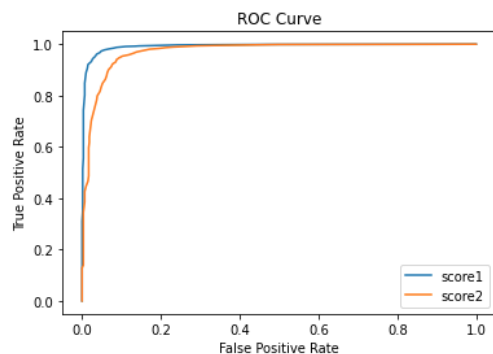
Ideal classifier model should be the model which completely distinguish between true class and false class. So, there will be no false positive and false negative values and the ROC curve will be like below



And the AUC under the curve will be the maximum.

We can decide the model is working better based on ROC curve. Of which model ROC is closer to the ROC of an ideal model we can conclude that model is working better or classifying better.

Here for our case, there are two models of which we can see that the Model1 ROC is closer to the ROC of ideal model hence we can conclude that the Model1 is a better classifier.



3. Compute AUC for both the ROC curves. (5 marks)

4. Calculate true positive rate for both models when false acceptance rate is 10%

From the earlier explanation we get the model ROC more likely to the ideal ROC, will be the better model and from the graph itself we can see the area under the curve for Model1 is greater than that of the Model2 as Model1 is better classifier.

AUC:

Model1: 0.9923660388862101

Model2: 0.9718799286104365

Also, the Model1 curve will reach its saturation earlier to that of Model2.

False acceptance rate is nothing but the false positive rate. We will get higher true positive rate for a fixed false positive rate is that graph which reaches its saturation earlier.

Of our two model we can see that the Model1 curve is reaching its saturation earlier. So, for Model1 we will get higher TPR for same FPR that Model2.

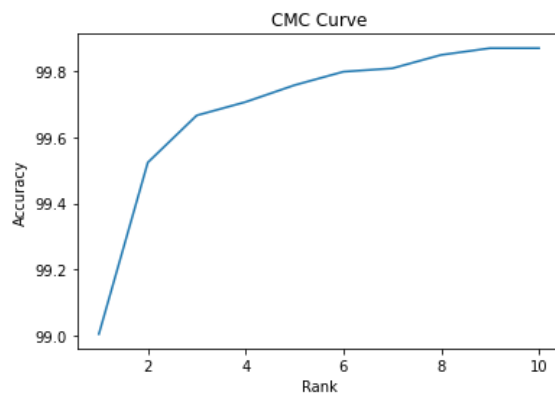
Same is being reflected in the result as well

TPR when FPR is 10:

Model1: 98.93%

Model2: 95.15%

Question 2



From the graph itself we can see that the accuracy is increasing with increasing in rank. Because here label we are getting a 2d array, of which each row only one value is 1 and rest are 0. And here we are comparing row-wise if score array holds the same value as label.

With increase in rank, column dimension of score sub-matrix is increasing as a result possibility of matching with label is also increasing. Which can be seen in the graph.

In the graph we can also see the graph flattens at Rank 6-7. Means there is almost negligible change in accuracy with increase in Rank. Whereas with increase in rank we are increasing in computational cost. So, we can take till rank 6 as well, which will hardly cost towards accuracy but we can reduce computational cost.