Kshitija Shelar

Text Analytics COMP4760

Date: 21/11/20

20200155

Practical 8

1. Confusion matrix: This is a statistical table which is used to evaluate the performance of machine learning algorithm. It contains important factors which can show how well the algorithm has performed. It has got 4 different values of predicted and actual values.

Actual \ Predicted		
	TP	FN
	FP	TN

TP (True Positive): when correct prediction of positive class happens, i.e when the actual value is 1 and the predicted value is also 1.

TN (True Negative): When correct prediction of negative class happens, i.e when the actual value is 0 and the predicted value is also 0.

FP (False Positive): When incorrect prediction of positive class happens, i.e when the actual value is 0 and predicted value is 1.

FN (False Negative): When incorrect prediction of negative class happens, i.e when the actual value is 1 and predicted value is 0.

Precision: It is the amount/fraction of relevant output among the retrieved outputs.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall: It is the amount/fraction of relevant output that are actually retrieved.

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1 score: it is used to measure accuracy of model using precision and recall. It is calculated when more importance is to be given on FPs and FNs.

$$F1 = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

False Positive Rate: it is proportion of false positives by total number of actual negatives.

$$FPR = \frac{FP}{FP + TN}$$

False negative rate: proportion of known positives for which result is negative.

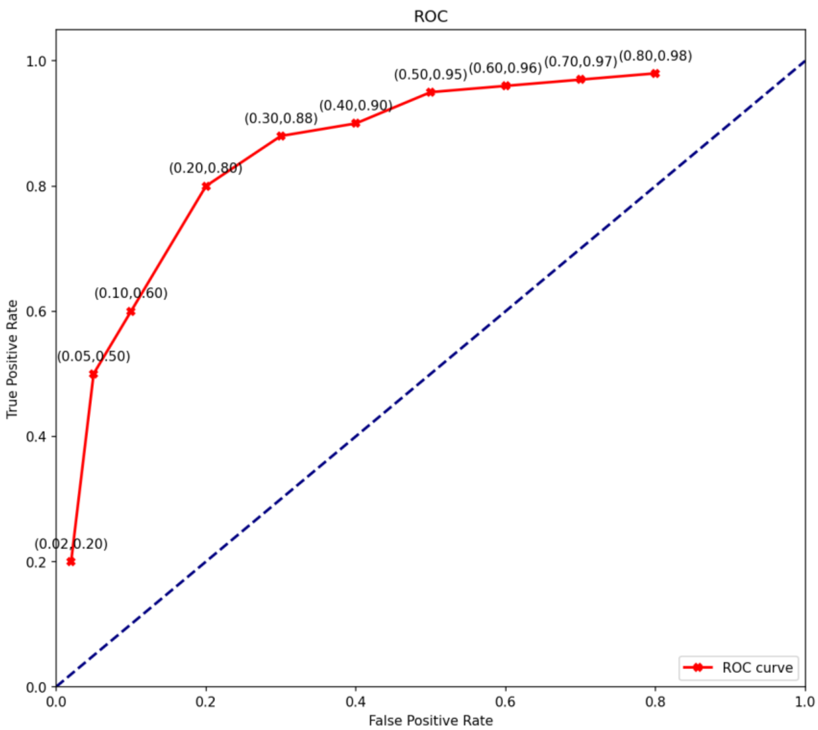
$$FNR = \frac{FN}{FN + TP}$$

For the data given, I calculated Precision, Recall, FPR, FNR and F1 score.

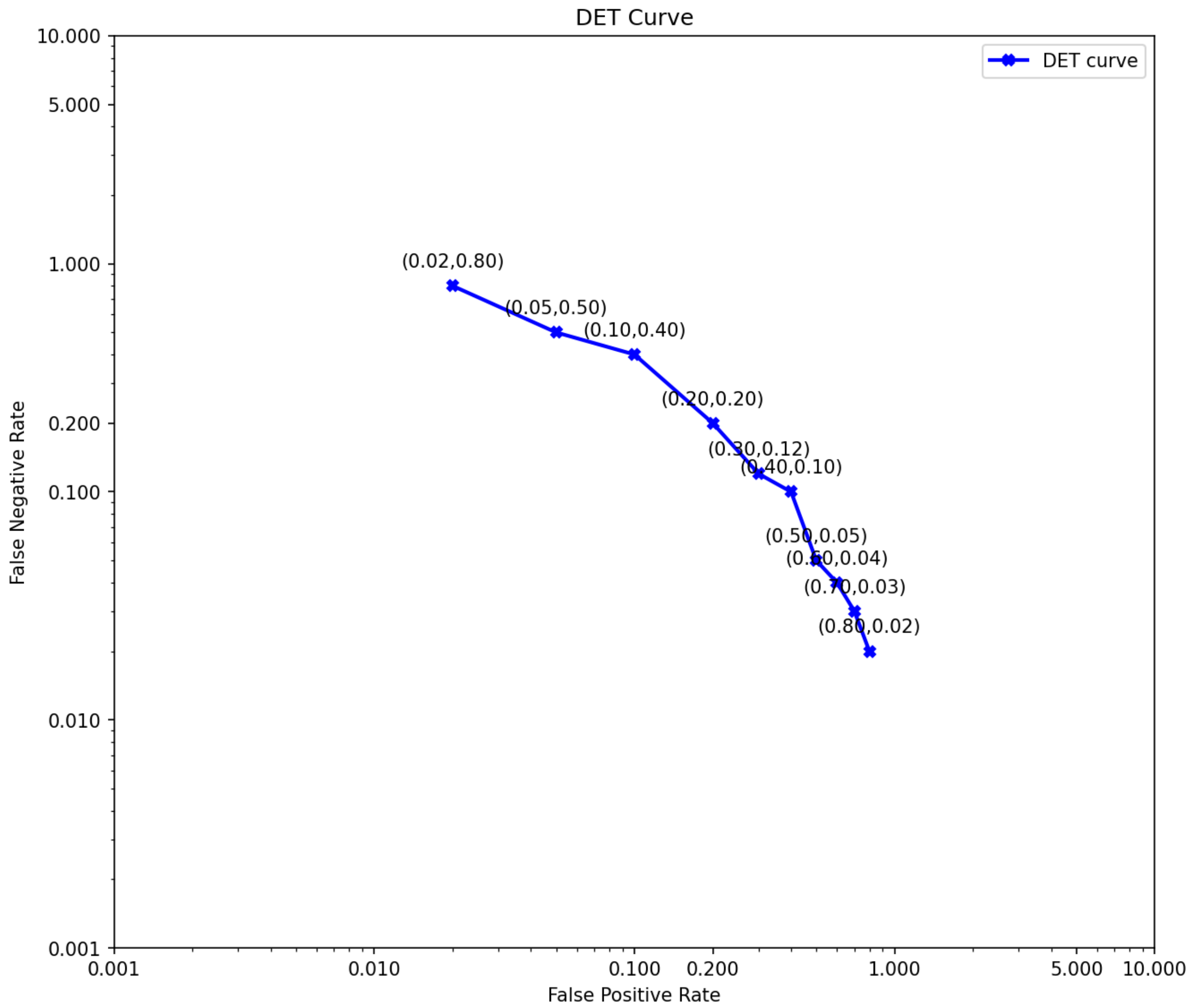
Index	Threshold	TP	FN	FP	TN	Correct	Incorrect	test Set	Precision	Recall/TPR	FPR	FNR/Miss-Detection	F1-Measure
0	1	20	80	2	98	100	100	200	0.909091	0.2	0.02	0.8	0.327869
1	5	50	50	5	95	100	100	200	0.909091	0.5	0.05	0.5	0.645161
2	10	60	40	10	90	100	100	200	0.857143	0.6	0.1	0.4	0.705882
3	15	80	20	20	80	100	100	200	0.8	0.8	0.2	0.2	0.8
4	20	88	12	30	70	100	100	200	0.745763	0.88	0.3	0.12	0.807339
5	25	90	10	40	60	100	100	200	0.692308	0.9	0.4	0.1	0.782609
6	30	95	5	50	50	100	100	200	0.655172	0.95	0.5	0.05	0.77551
7	35	96	4	60	40	100	100	200	0.615385	0.96	0.6	0.04	0.75
8	40	97	3	70	30	100	100	200	0.580838	0.97	0.7	0.03	0.726592
9	50	98	2	80	20	100	100	200	0.550562	0.98	0.8	0.02	0.705036

We can observe that for threshold of 20, the F1 measure is the highest i.e 0.807.

2. ROC:
It is used for measuring performance mainly of classification tasks using different thresholds [1]. "ROC is created by plotting the TPR against FPR at various threshold settings" [2].
AUC is Area Under the ROC Curve. It represents measure of separability [1]. For higher AUC, performance of that model is better. i.e more the area under the curve, higher is the accuracy of that particular model.



3. DET curve:
"It is a graphical plot of error rates for binary classification systems, plotting the false rejection rate vs false acceptance rate. The x-and y- axes are scaled non-linearly by their standard normal derivatives (or just by logarithmic transformation), yielding trade off curves that are more linear than ROC curves, and use most of the image area to highlight the differences of importance in the critical operating region".[3]



References:

[1]https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5

[2] https://en.wikipedia.org/wiki/Receiver_operating_characteristic

[3] https://en.wikipedia.org/wiki/Detection_error_tradeoff