Data Application Lab

Quiz for Machine Learning Basis II Answer

Problem 1

a) Label A – F

A	False Negative
B	True Negative
C	True Positive
D	False Positive
E	Recall
F	Precision

b) Use A – D to represent quantities, Accuracy and false Positive Rate

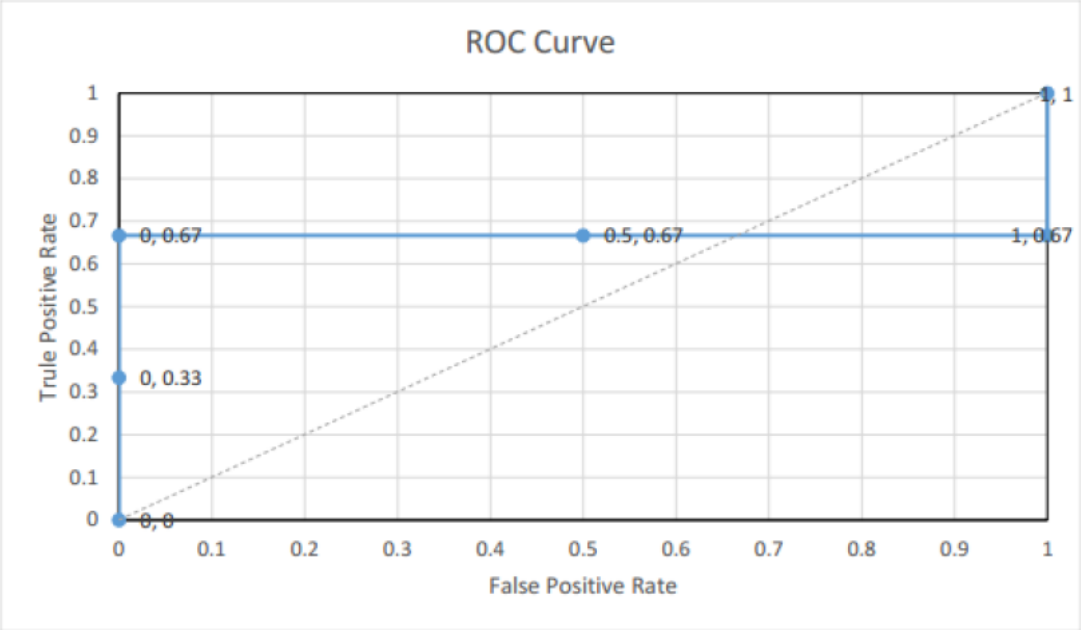
$$\text{Accuracy} = \frac{True\ Positive + True\ Negative}{Total\ Population} = \frac{C + B}{A + B + C + D}$$

$$False\ \text{Positive Rate} = \frac{False\ Positive}{Actual\ Negative} = \frac{False\ Positive}{False\ Positive + True\ Negative} = \frac{D}{D+B}$$

Problem 2

a) Manually plot the ROC curve

Classifier Output	True Class	Threshold 0.2	Threshold 0.3	Threshold 0.5	Threshold 0.6	Threshold 0.8	Threshold 0.9
0.24	1	T	F	F	F	F	F
0.45	0	T	T	F	F	F	F
0.57	0	T	T	T	F	F	F
0.73	1	T	T	T	T	F	F
0.89	1	T	T	T	T	T	F
TPR		3/3 = 1	2/3	2/3	2/3	1/3	0/3 = 0
FPR		2/2 = 1	2/2 = 1	1/2 = 0.5	0/2 = 0	0/2 = 0	0/2 = 0



b) Calculated the AUC of the above ROC curve

Area under the ROC curve above: $1 * 2/3 = 2/3 \approx 0.67$

c) In general, why should AUC >= 0.5?

A worthless test, which does not discriminate between class 0 and 1, would have a curve as the diagonal line (dotted grey line). As the diagonal line bisects the graph, the area under ROC curve for a worthless test is 0.5. Therefore, area under ROC curve is greater or equal to 0.5 in general.