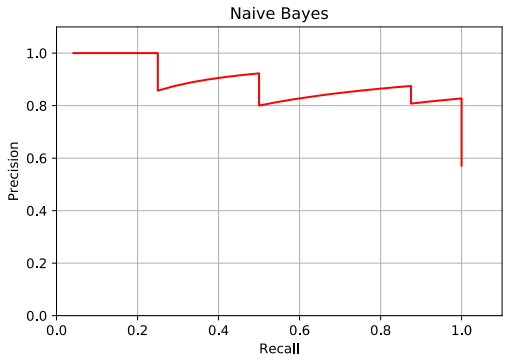
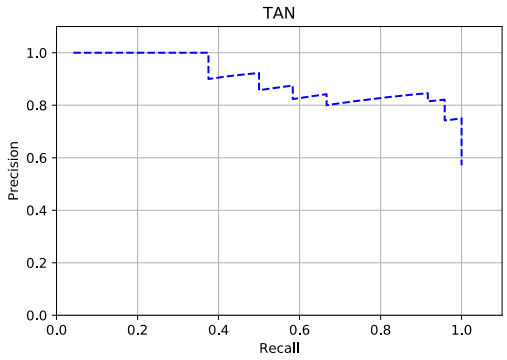
Part 2 –

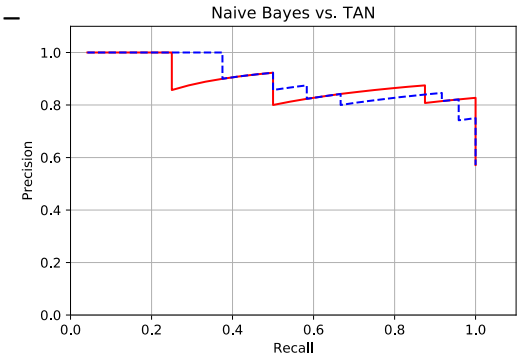
1. Naïve Bayes –



2. TAN –



3. Naïve Bayes vs TAN –



TAN seems to have more predictive power since the area under the curve is greater than in the case of Naïve Bayes.

In datasets where the baseline prior probability is important, i.e., when one of the class is much rarer than the other or if one class is much more interesting than the other, PR curves give more insight about the algorithm than ROC curves since they are influenced by 'true negatives' which might not be an important metric if the positive class is more important or much more rare. In case of cancer, 'metastases' class is much more important and hence a PR curve is a better measure for this dataset.

Part 3 –

Fold 1 –
Naïve Bayes accuracy = 0.9333333333333333
TAN accuracy = 0.8666666666666667
Accuracy delta = -0.06666666666

Fold 2 –
Naïve Bayes accuracy = 0.7333333333333333
TAN accuracy = 0.8666666666666667
Accuracy delta = 0.13333333333

Fold 3 –
Naïve Bayes accuracy = 0.9285714285714286
TAN accuracy = 0.9285714285714286
Accuracy delta = 0

Fold 4 –
Naïve Bayes accuracy = 0.8571428571428571
TAN accuracy = 0.9285714285714286
Accuracy delta = 0.07142857142

Fold 5 –
Naïve Bayes accuracy = 0.7142857142857143
TAN accuracy = 0.9285714285714286
Accuracy delta = 0.21428571428

Fold 6 –
Naïve Bayes accuracy = 0.7142857142857143
TAN accuracy = 0.7142857142857143
Accuracy delta = 0

Fold 7 –
Naïve Bayes accuracy = 0.9285714285714286
TAN accuracy = 1.0
Accuracy delta = 0.07142857142

Fold 8 –
Naïve Bayes accuracy = 0.7857142857142857
TAN accuracy = 0.8571428571428571
Accuracy delta = 0.07142857142

Fold 9 –
Naïve Bayes accuracy = 0.8571428571428571
TAN accuracy = 0.8571428571428571
Accuracy delta = 0

Fold 10 –
Naïve Bayes accuracy = 0.7142857142857143
TAN accuracy = 0.7142857142857143
Accuracy delta = 0

Naïve Bayes mean accuracy = 0.8166666666666667
TAN mean accuracy = 0.8661904761904762
Mean accuracy delta = 0.049523809521

t-statistic $= \frac{0.0495}{\sqrt{\frac{0.0584}{90}}} =$ 1.9355

Looking up the t-value from the table with df = 9, we have t = 2.262.
P-value = 0.0849. Hence, it is not significant.