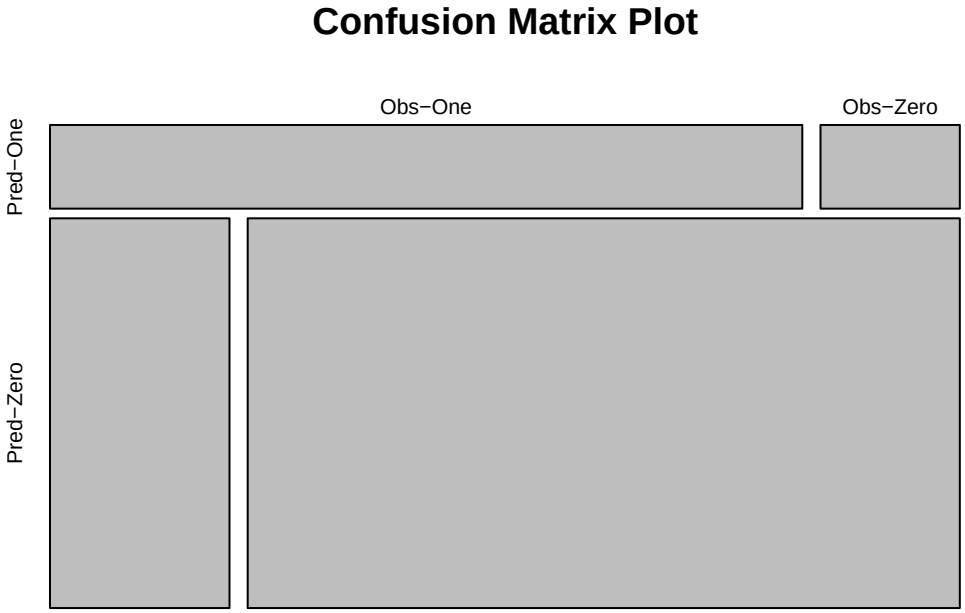
IS621_Assignment2

Senthil Dhanapal

October 6, 2016

Confusion Matrix

```
##
##              Obs-One Obs-Zero
##   Pred-One        27        5
##   Pred-Zero       30      119
```



5. Write a function that takes the data set as a dataframe, with actual and predicted classifications identified, and returns the precision of the predictions.

Precision = 0.84375

6. Write a function that takes the data set as a dataframe, with actual and predicted classifications identified, and returns the sensitivity of the predictions. Sensitivity is also known as recall.

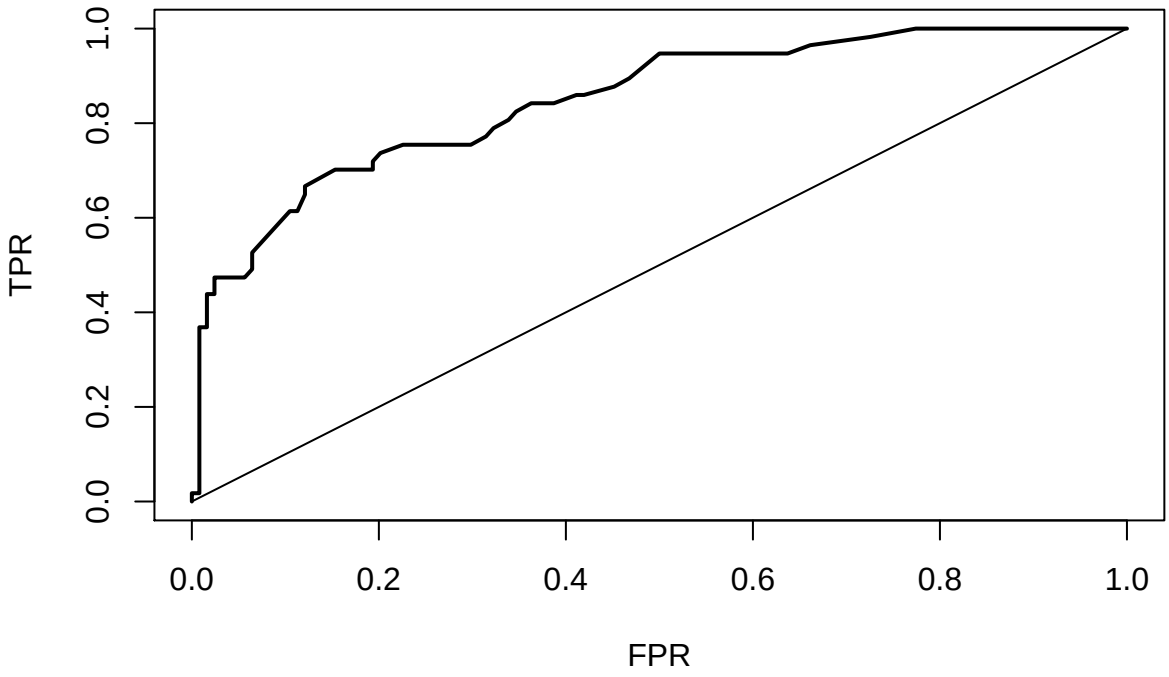
Sensitivity = 0.4736842

10. Write a function that generates an ROC curve from a data set with a true classification column (class in our example) and a probability column (scored.probability in our example). Your function should return a list that includes the plot of the ROC curve and a vector that contains the calculated area under the curve (AUC). Note that I recommend using a sequence of thresholds ranging from 0 to 1 at 0.01 intervals.

I calculated best threshold value using two methods.
a) by calculating the distance of the point from (0,1).
b) by calculating AUC by making a curve for threshold{t} by joining points (0,0), (X{t},Y{t}), (1,1)

I used two thresholds increments - 1) .01, 2) .001

- 1) when cut-off is threshold(0,1,0.01) -> value is very close to what pROC predicts

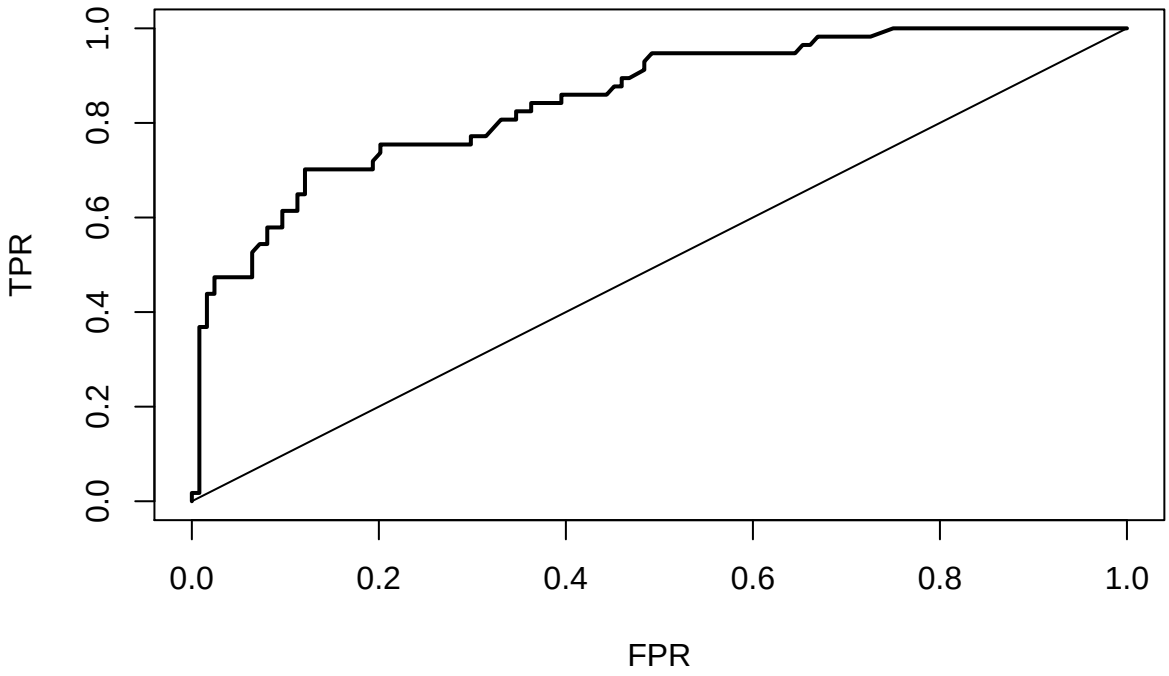


AUC using manual calculation is 0.8488964.

Best Threshold value using method 1 is {Threshold = 0.320000,fpr = 0.201613,tpr = 0.736842,auc = 0.767615, dist = 0.331511}

Best Threshold value using method 2 is {Threshold = 0.370000,fpr = 0.153226,tpr = 0.701754,auc = 0.774264, dist = 0.335304}

- 2) when cut-off is threshold(0,1,0.001) -> value is exact to what pROC predicts

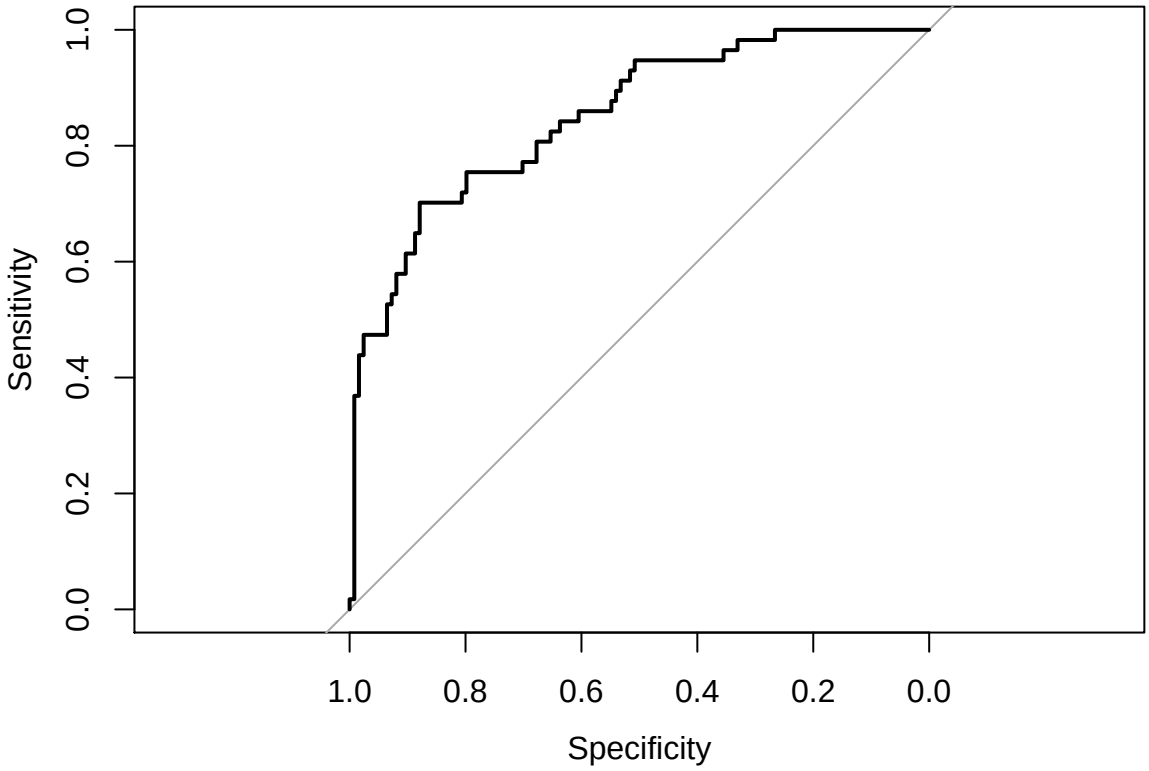


AUC using manual calculation is 0.850382.

Best Threshold value using method 1 is {Threshold = 0.316000,fpr = 0.201613,tpr = 0.754386,auc = 0.776387, dist = 0.317764}

Best Threshold value using method 2 is {Threshold = 0.374000,fpr = 0.120968,tpr = 0.701754,auc = 0.790393, dist = 0.321844}

13. Investigate the pROC package. Use it to generate an ROC curve for the data set. How do the results compare with your own functions?




```
##
## Call:
## roc.default(response = dfData$class, predictor = dfData$scored.probability)
##
## Data: dfData$scored.probability in 124 controls (dfData$class 0) < 57 cases (dfData$class 1).
## Area under the curve: 0.8503
```

Best Threshold value using pROC package is {Threshold = 0.375117,fpr = 0.120968,tpr = 0.701754}

Note: My second method (using auc) predicts better than first method (using distance from (0,1))