9. Confusion matrix problem

Solutions for a)::

Section 1

True negative is 50

False positive is 10

False negative is 5

True positive is 100

Secton 2:

(TP+TN+FP+FN)=n

In our case n=165

Accuracy = (TP+TN)/(TP+TN+FP+FN)

Accuracy1 = (100+50)/(100+50+5+10)=150/165 = 0.9(09)

Section 3:

Sensitivity=TPR=TP/P=TP/(TP+FN)

Sensitivity1=100/(100+5) approximately equal to 0.95

Specificity=TNR=TN/(TN+FP)

Specificity1=50/(50+5)= 0.9(09)   
False Positive Rate = FP/(FP+TN)

False Positive Rate 1= 10/(10+50)=⅙

Section 4:

Precision = PPV= TP/(TP+FP)= 50/(50+10)

Recall = Sensitivity As SEN AND TPR calculate the same thing but has different name so

Recall is approximately equal to 0.95

F-score = 2\*precision\*recall/(precision +Recall)= 2TP/( 2TP+FP+FN )= 2\*50/(2\*50+10+5)=approximately equal to 0.86

Solution to problem b)

K fold is an alternative method for LOOCV(Leave One-Out Cross Validation)

This approach involves randomly dividing the set of observations into k groups, or folds

,of approximately equal size.First told is treated as a validation set and the other k-1folds is to fit our data.Typically most people use k=5 or k=10 has computational advantage rather than mathematical one.Although k=n recursion might enable us to perform better statistical model however in practice it's not really recommended.By using another to subset to perform test and train again and since the datas are independent this will enable us to see if our model has been overfitting given one of the subsets which we performed cross validation.And we can iterate procedure over and over.Until we reached goal iteration (n).