EECS 658

Alex Anderson

NBClassifier Question Answers

a. There are 150 samples total samples and half of the total are used for training, so 75 samples are in the training set.
b. There are 150 samples total samples and half of the total are used for testing, so 75 samples are in the testing set.

Confusion Matrix for Questions c-f:

50	0	0
0	48	2
0	3	47

Formulas Used:

Acc = (sum of diagonal)/(total samples)
P = correct/(sum of row)
R = correct/(sum of column)
F1 = $2\cdot(P\cdot R)/(P+R)$

c. Total Accuracy = (50+48+47) / (150) = 0.966
    a. This value matches the accuracy calculated by the program.
d. Precision Values:
    a. P(Setosa)=50/50=100%
    b. P(Versicolor)=48/50=96%
    c. P(Virginica)=47/50=94%
        i. These values do not match the values calculated by the program. The precision values for Versicolor and Virginica are flipped.
e. Recall Values:
    a. R(Setosa)=50/50=100%
    b. R(Versicolor)=48/51=94%
    c. R(Virginica)=47/49=96%
        i. Similarly, the recall values for Versicolor and Virginica are flipped.
f. F1 Values:
    a. F1(Setosa)=2((1*1)/ (1+1)) =100%
    b. F1 (Versicolor)=2((.96*.94)/ (.96+.94)) =95%
    c. F1 (Virginica)=2((.94*.96)/ (.94+.96)) =95%
        i. These values do match since the flipping of precision and recall values was symmetrical.