Homework02

Hwijong Im

Problem01

(1) False Positive(FP) = 3, False Negative(FN) = 1

(2) Accuracy = (TP+TN) / (TP+FP+FN+TN) = 105/109 = 0.96

(3) Pricision = TP / (TP + FP) = 33 / (33+3) = 11/12 = 0.92
Recall = TP / (TP+FN) = 33/34 = 0.97

(4) F1 Score = 2 / (1/Precision)+(1/Recall) = 66/(36+34) = 66/70 = 0.94

(5) I would take higher recall. The crucial point of COVID test is that the tester must search out people who have actually got positive reaction in COVID test. Therefore, the model should have high performance to decrease False Negative.

Problem03

Manhattan Distance = |X(f1) - X_new(f1)| + |X(f2) - X_new(f2)| + |X(f3) - X_new(f3)|

	f1	f2	f3	y	Manhattan Distance(d)
x1	1	4	1	1	4
x2	1	2	3	1	4
x3	0	0	1	1	1
x4	-1	4	0	1	7
x5	1	0	-2	0	3
x6	-1	-1	1	0	3
x7	0	-4	0	0	6
x8	1	0	3	0	2
x_new	1	0	1		

k=3 Nearest neighbors based on the Manhattan Distance

k=3	f1	f2	f3	y	Manhattan Distance (d)
x3	0	0	1	1	1
x8	1	0	-3	0	2
x5	1	0	-2	0	3
x6	-1	-1	1	0	3

Therefore, based on x8, x6, and x6, the x_new is classified as y=0 in k=3 Manhattan Distance

Problem04

Naive Bayes Assumption = P(X|Y)*P(Y) / P(X)

P(Y=Yes) = 9/14 P(Y=No) = 5/14

P(Y=Yes|X=Youth, Medium, Yes, Fair) = P(X|Y=Yes)*P(Y=Yes)
= 2/9 * 4/9 * 6/9 * 6/9 = 288 / 6561

P(Y=No|X=Youth, Medium, Yes, Fair) = P(X|Y=No)*P(Y=No)
= 3/5 * 2/5 * 1/5 * 2/5 = 12 / 625

P(X) = P(X|Y=Yes) + P(X|Y=No) = 0.03507

Therefore, P(Y=Yes|X) = 288/6561 / 0.03507 = 0.8043 > 0.5
Classification will be Y = Yes
This client will buy the computer