HW3
Deadline: 12/11

1. Consider the data set shown in Table 1.

Table 1. Data set.

Instance	A	B	C	Class
1	0	0	1	$-$
2	1	0	1	$+$
3	0	1	0	$-$
4	1	0	0	$-$
5	1	0	1	$+$
6	0	0	1	$+$
7	1	1	0	$-$
8	0	0	0	$-$
9	0	1	0	$+$
10	1	1	1	$+$

a) (18%) Estimate the conditional probabilities for P(A = 1|+), P(B = 1|+), P(C = 1|+), P(A = 1|−), P(B = 1|−), and P(C = 1|−).

P(A = 1|+):
Class = + 且 A = 1：3
Class = +：5
P(A = 1|+) = Count(A = 1 且 Class = +) / Count(Class = +) = 3/5

P(B = 1|+):
Class = + 且 B = 1：2
Class = +：5
P(B = 1|+) = Count(B = 1 且 Class = +) / Count(Class = +) = 2/5

P(C = 1|+):
Class = + 且 C = 1：4
Class = +：5
P(C = 1|+) = Count(C = 1 且 Class = +) / Count(Class = +) = 4/5

P(A = 1|-):
Class = - 且 A = 1：2
Class = -：5
P(A = 1|-) = Count(A = 1 且 Class = -) / Count(Class = -) = 2/5

P(B = 1|-):
Class = - 且 B = 1：2
Class = -：5
P(B = 1|-) = Count(B = 1 且 Class = -) / Count(Class = -) = 2/5

P(C = 1|-):
Class = - 且 C = 1：1
Class = -：5
P(C = 1|-) = Count(C = 1 且 Class = -) / Count(Class = -) = 1/5

b) (10%) Use the conditional probabilities in part (a) to predict the class label for a test sample (A = 1, B = 1, C = 1) using the naïve Bayes approach.

Let R: (A=1, B=1, C=1) be the test record.
To determine its class, we need to compute P(+|R) and P(−|R). Using Bayes theorem,

P(+|R) =P(R|+) P(+) / P(R) and P(−|R) =P(R|−) P(−) / P(R).
Since P(+) = P(−) = 0.5 and P(R) is constant,
R can be classified by comparing P(+|R) and P(−|R). For this question,

P(R|+) = P(A=1|+) × P(B=1|+) × P(C=1|+) = 0.192
P(R|−) = P(A=1|−) × P(B=1|−) × P(C=1|−) = 0.032

Since P(R|+) is larger, the record is assigned to (+) class.

2. Consider the one-dimensional data set shown in Table 2.

Table 2. Data set.

x	0.5	3.0	4.5	4.6	4.9	5.2	5.3	5.5	7.0	9.5
y	−	−	+	+	+	−	−	+	−	−

a) (12%) Classify the data point x = 5.0 according to its 1-, 3-, 5-, and 9-nearest neighbors using **majority voting**.

多數投票法（majority Voting）：
1- 最近鄰：
x = 5.0 的 1- 最近鄰是 {4.9}，有1個 (+)，所以 x = 5.0 會被歸類為 (+)。
3- 最近鄰：
x = 5.0 的 3- 最近鄰是 {4.9, 5.2, 5.3}，其中有1個 (+) 和2個 (-)，所以 x = 5.0 會被歸類為 (-)。
5- 最近鄰：
x = 5.0 的 5- 最近鄰是 {4.5, 4.6, 4.9, 5.2, 5.3}，其中有3個 (+) 和2個 (-)，所以 x = 5.0 會被歸類為 (+)。
9- 最近鄰：
x = 5.0 的 9- 最近鄰是 {0.5, 3.0, 4.5, 4.6, 4.9, 5.2, 5.3, 5.5, 7.0}，其中有 4 個 (+) 和 5 個 (-)，所以 x = 5.0 會被歸類為 (-)。

b) (12%) Classify the data point x = 5.0 according to its 1-, 3-, 5-, and 9-nearest neighbors using the **distance-weighted voting** approach.

距離加權投票法（distance-weighted voting）：
1- 最近鄰：
x = 5.0 的 1- 最近鄰是 {4.9}，有1個 (+) ，Distance: |4.9 - 5.0| = 0.1，所以 x = 5.0 會被歸類為 (+)。
3- 最近鄰：
x = 5.0 的 3- 最近鄰是 {4.9, 5.2, 5.3}，其中有1個 (+) 和2個 (-)，Distance: **|4.9 - 5.0| = 0.1**, |5.2 - 5.0| = 0.2, |5.3 - 5.0| = 0.3，所以 x = 5.0 會被歸類為 (+)。
5- 最近鄰：
x = 5.0 的 5- 最近鄰是 {4.5, 4.6, 4.9, 5.2, 5.3}，其中有3個 (+) 和2個 (-)，Distance: |4.5 - 5.0| = 0.5, |4.6 - 5.0| = 0.4, **|4.9 - 5.0| = 0.1**, |5.2 - 5.0| = 0.2, |5.3 - 5.0| = 0.3，所以 x = 5.0 會被歸類為 (+)。
9- 最近鄰：
x = 5.0 的 9- 最近鄰是 {0.5, 3.0, 4.5, 4.6, 4.9, 5.2, 5.3, 5.5, 7.0}，其中有 4 個 (+) 和 5 個 (-)，Distance: |0.5 - 5.0| = 4.5, |3.0 - 5.0| = 2.0, |4.5 - 5.0| = 0.5, |4.6 - 5.0| = 0.4, **|4.9 - 5.0| = 0.1**, |5.2 - 5.0| = 0.2, |5.3 - 5.0| = 0.3, |7.0 - 5.0| = 2.0，所以 x = 5.0 會被歸類為 (+)。

3. You are asked to evaluate the performance of two classification models, M1 and M2. The test set you have chosen contains 26 binary attributes, labeled as A through Z.

   Table 3 shows the posterior probabilities obtained by applying the models to the test set. (Only the posterior probabilities for the positive class are shown). As this is a two-class problem, P(−) = 1−P(+) and P(−|A,...,Z) = 1−P(+|A,...,Z). Assume that we are mostly interested in detecting instances from the positive class.

Table 3. Posterior probabilities.

Instance	True Class	$P(+\|A,\ldots,Z,M_1)$	$P(+\|A,\ldots,Z,M_2)$
1	+	0.73	0.61
2	+	0.69	0.03
3	−	0.44	0.68
4	−	0.55	0.31
5	+	0.67	0.45
6	+	0.47	0.09
7	−	0.08	0.38
8	−	0.15	0.05
9	+	0.45	0.01
10	−	0.35	0.04

a) (12%,3%) Plot the ROC curve for both M1 and M2. (You should plot them on the same graph.) Which model do you think is better? Explain your reasons.

The ROC curve for M1 and M2 are shown in the Figure 4.5. M1 is better, since its area under the ROC curve is larger than the area under ROC curve for M2.

b) (15%) For model M1, suppose you choose the cutoff threshold to be t = 0.5. In other words, any test instances whose posterior probability is greater than t will be classified as a positive example. Compute the precision, recall, and F-measure for the model at this threshold value.

Precision=3/4=75%. Recall=3/5=60%

F-measure=(2×.75×.6)/(.75+.6)=0.667.

c) (15%,2%,1%) For model M2, suppose you choose the cutoff threshold to be t = 0.5. Compute the precision, recall, and F-measure for model M2 at this threshold value. Compare the F-measure results for both models. Which model is better? Are the results consistent with what you expect from the ROC curve?

Precision = 1/2 = 50%. Recall = 1/5 = 20%.

F-measure = (2×.5×.2) / (.5+.2) = 0.2857. Based on F-measure,

M1 is still better than M2. This result is consistent with the ROC plot.