[redacted]

[redacted]

1.

Consider the training dataset given below. A, B, and C are the attributes and Y is the class variable.

A	B	C	Y
0	1	0	Yes
1	0	1	Yes
0	0	0	No
1	0	1	No
0	1	1	No
1	1	0	Yes

(a) (2 points) Can you draw a decision tree having 100% accuracy on this training set? If you answer is yes, draw the decision tree in the space provided below. If your answer is no, explain why?

Solution: No. Because Examples #2 and #4 have the same feature vector but different classes.

(b) (3 points) Which attribute among A, B and C has the highest information gain? Explain your answer.

Solution:
$IG(A) = H(3,3) - \frac{1}{2}H(2,1) - \frac{1}{2}H(2,1)$
$IG(B) = H(3,3) - \frac{1}{2}H(2,1) - \frac{1}{2}H(2,1)$
$IG(C) = H(3,3) - \frac{1}{2}H(2,1) - \frac{1}{2}H(2,1)$
Therefore, all three attributes have the same information gain.

2.

The models we have been working with above are global models, in the sense that the parameters are the same, no matter where we are in the feature space.

(a) **In what sense is the KNN classifier a "local" model?**

The KNN classifier, for a given set of feature values x_0, uses only the K observations in the training set whose features are "most similar" to x_0; that is it uses only the "local" values. ("Most similar" usually means closest in a Euclidean distance sense, but it could mean something else.)

(b) Suppose we are given a training set in a binary classification problem, and now we want to classify a new observation with features x_0. **Describe how you would classify that observation.**

As described above...

[redacted]

[redacted]

3.

Consider the following training dataset with two real-valued inputs X_1 and X_2 and a class variable Y that takes two values, $+$ and $-$.

X_1	X_2	Y
0	5	$+$
2	9	$+$
1	3	$-$
2	4	$-$
3	5	$-$
4	6	$-$
5	7	$-$

We assume that the data is generated by a Gaussian naive Bayes model, and we will use the data to develop a naive Bayes classifier.

(a) (5 points) Assuming that the variance is independent of the class, estimate the parameters of the Gaussian Naive Bayes model from the given dataset.

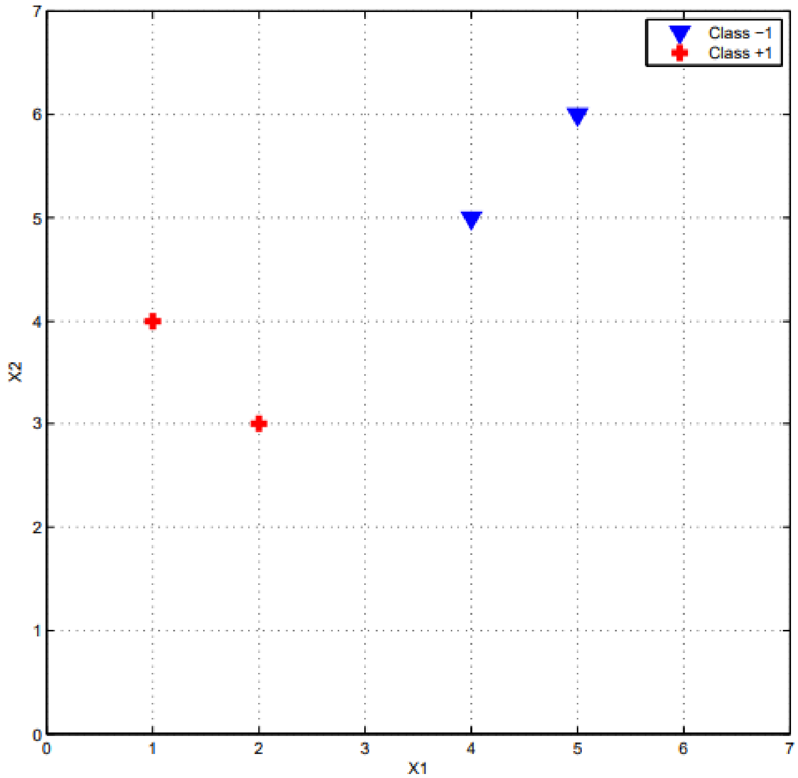
Solution: The parameters are $P(Y = +) = 2/7$ and $P(Y = -) = 5/7$ The mean of the Gaussian representing $P(X_1|Y = +)$ is $(0 + 2)/2 = 1$ and with $P(X_1|Y = -)$ is $(1 + 2 + 3 + 4 + 5)/5 = 3$. Since the variance is independent of the class, the variance associated with both Gaussians is the variance of the vector $(0, 2, 1, 2, 3, 4, 5)$. The mean of the Gaussian representing $P(X_2|Y = +)$ is $(5 + 9)/2 = 7$ and with $P(X_2|Y = -)$ is $(3 + 4 + 5 + 6 + 7)/5 = 5$. Since the variance is independent of the class, the variance associated with both Gaussians is the variance of the vector $(5, 9, 3, 4, 5, 6, 7)$.

(b) (5 points) Assuming that the variance is independent of the features X_1 and X_2, estimate the parameters of the Gaussian Naive Bayes model from the given dataset.

Solution: The parameters are $P(Y = +) = 2/7$ and $P(Y = -) = 5/7$ The mean of the Gaussian representing $P(X_1|Y = +)$ is $(0 + 2)/2 = 1$ and with $P(X_1|Y = -)$ is $(1 + 2 + 3 + 4 + 5)/5 = 3$. The mean of the Gaussian representing $P(X_2|Y = +)$ is $(5 + 9)/2 = 7$ and with $P(X_2|Y = -)$ is $(3 + 4 + 5 + 6 + 7)/5 = 5$.

Since the variance is independent of the features, the variance associated with Gaussians representing $P(X_1|Y = +)$ and $P(X_2|Y = +)$ is the variance of the vector $(0, 2, 5, 9)$ while the variance associated with Gaussians $P(X_1|Y = -)$ and $P(X_2|Y = -)$ is the variance of the vector $(1, 2, 3, 4, 5, 3, 4, 5, 6, 7)$.

4.



Support vector machines learn a decision boundary leading to the largest margin from both classes. You are training SVM on a tiny dataset with 4 points shown in Figure 2. This dataset consists of two examples with class label -1 (denoted with plus), and two examples with class label +1 (denoted with triangles).

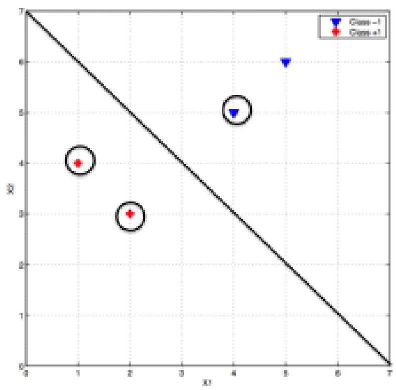
i. Find the weight vector **w** and bias b. What's the equation corresponding to the decision boundary?

Solution:
SVM tries to maximize the margin between two classes. Therefore, the optimal decision boundary is diagonal and it crosses the point (3,4). It is perpendicular to the line between support vectors (4,5) and (2,3), hence it is slope is m = -1. Thus the line equation is $(x_2 - 4) = -1(x_1 - 3) = x_1 + x_2 = 7$. From this equation, we can deduce that the weight vector has to be of the form (w_1, w_2), where $w_1 = w_2$. It also has to satisfy the following equations:
$2w_1 + 3w_2 + b = 1$ and
$4w_1 + 5w_2 + b = -1$

Hence $w_1 = w_2 = -1/2$ and $b = 7/2$

. Circle the support vectors and draw the decision boundary.

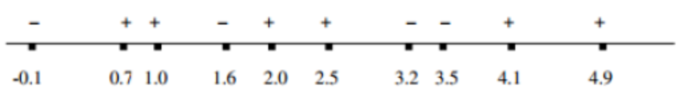
Solution:



5.

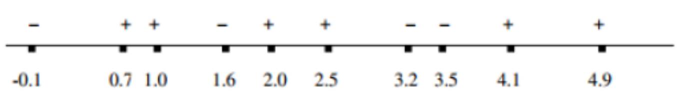
Problem 12 Consider the following dataset with one real-valued input and one binary output (+ or -). The following questions assume that we are using k-nearest-neighbor learning with unweighted Euclidean distance to predict y for an input x.

X	Y
-0.1	-
0.7	+
1.0	+
1.6	-
2.0	+
2.5	+
3.2	-
3.5	-
4.1	+
4.9	+



A. (2 points) What is the leave-one-out cross-validation error of 1-NN on this dataset. Give your answer as the number of misclassifications and circle them in the diagram above.
4. For each X, if the nearest neighbor has a different label, X would be misclassified.

B. (2 points) What is the leave-one-out cross-validation error of 3-NN on this dataset. Give your answer as the number of misclassifications and circle them in the diagram below.



8. For each X, consider the majority vote of 3 nearest neighbors.