Sheet 3

Q1

Consider the following 2-dimensional dataset. We have 8 points and 2 classes {+, -}

Feature 1	Feature 2	Class
-1	0	+
-0.5	0.5	+
0	1	-
0.5	1	-
1	0	+
1	-1	+
0	-1	-
0	0	-

In this problem, you will run 3 iterations of AdaBoost with decision stumps (one-level decision trees) as weak learners.

(a) Plot the data

(b) For each iteration:

- Compute the weight for each data point
- Sketch the decision boundary
- Compute the weighted error
- Compute the model (learner) weight

(c) Compute the training error after the 3 iterations

Q2

In the following data set:

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

Use Naive Bayes classification to classify (Red, SUV, Domestic).

Q3

Consider the data set shown in the following table:

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

(a) Estimate the conditional probabilities for P(A | +), P(B | +), P(C | +), P(A | -), P(B | -), and P(C | -).

(b) Use the estimate of conditional probabilities given in the previous question to predict the class label for a test sample (A = 0, B = 1, C = 0) using the Naive Bayes approach.

Q4

Given the Data Matrix on the right answer the following questions.

a. What is number of dimensions?
b. What are the types of the attributes?
c. What is the distance between x1 and x3?
d. What is the length of x2?
e. What is the cos(angle) between x2 and x4?

ID	a1	a2	a3	a4
1	10	60	10	90
2	20	50	40	70
3	30	50	30	40
4	20	50	20	60
5	10	60	30	10

DATA MATRIX **D**

Q5

Given Data matrix above. Consider a1, a2 and a4 only.

a. Write down the new data matrix **D3** (5x3).
b. Plot the data using 3d scatter plots.
c. Compute the mean vector (3x1).
d. Compute centered data matrix **Z** by subtracting mean vector from the Data Matrix. (5x3).
e. Compute Covariance matrix **COV** (3x3).