Homework 11: finish by 6/24

Reading: Notes: Chapter 10

Videos: 10.1 - 10.3

Problem 11.1 (**Video 10.1, 10.2, 10.3**, **Lecture Problem**)

You are given the training data and testing data in the tables below.

Training Data

x_1	x_2	label
1	3	+
3	3	+
2	1	+
4	1	+
-1	-3	-
-3	-3	-
-2	-1	-
-4	-1	-

Testing Data

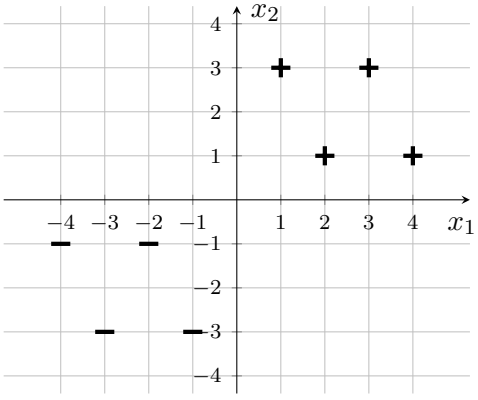
x_1	x_2	label
1	2	+
1	-3	+
-2	-3	-
-2	1	-

(a) Make a two-dimensional plot of the training data (using pluses and minuses). Determine the sample mean vectors $\underline{\mu}_+$ and $\underline{\mu}_-$ and add them to the plot.

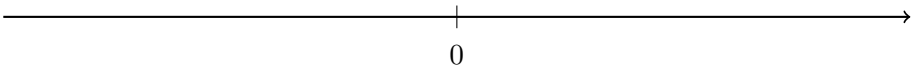
(b) On your plot, sketch the decision boundary for the closest average classifier.

(c) Determine the number of training errors and the number of test errors for the closest average classifier.

(d) Sketch the Gaussian contour plot for PCA dimensionality reduction applied to the entire training dataset. Draw the eigenvector that would be used to project down from two dimensions to one dimension.

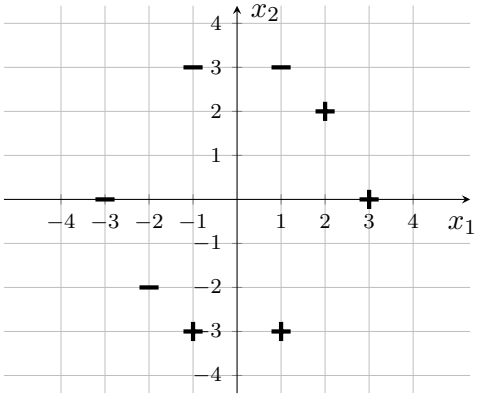


(e) On a one-dimensional axis such as the one below, sketch the pluses and minuses for the training data after PCA dimensionality reduction to a single dimension, using your vector from (d). (You do not need to calculate the exact values for each one-dimensional point.) Sketch the decision boundary for the resulting closest average classifier in one dimension.



Problem 11.2 (Video 10.1, 10.2, 10.3)

You are given the training data on the figure and table and the testing data in the table.



Training Data

x_1	x_2	label
2	2	+
3	0	+
1	-3	+
-1	-3	+
-2	-2	-
-3	0	-
-1	3	-
1	3	-

Testing Data

x_1	x_2	label
1	0	-
-1	0	-
0	1	-
0	-1	+

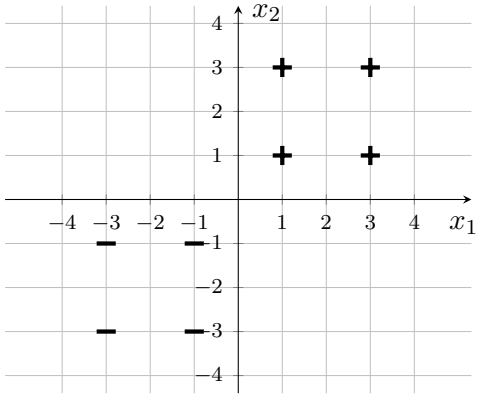
(a) Which of the following linear classifiers has the lowest training error?

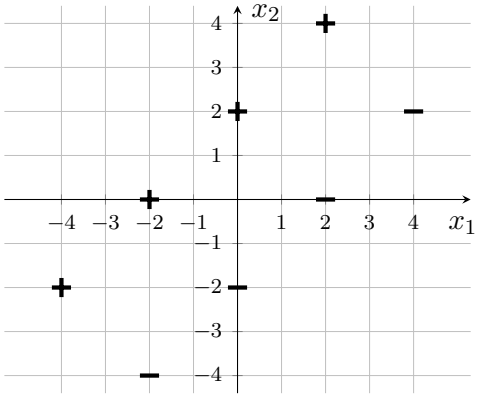
$$D_1(x_1,x_2)=\begin{cases}+ & 2x_2 \geq x_1\\ - & \text{otherwise}\end{cases} \quad D_2(x_1,x_2)=\begin{cases}+ & 2x_2 \leq x_1\\ - & \text{otherwise}\end{cases}$$

$$D_3(x_1,x_2)=\begin{cases}+ & x_2 \geq 2x_1\\ - & \text{otherwise}\end{cases} \quad D_4(x_1,x_2)=\begin{cases}+ & x_2 \leq 2x_1\\ - & \text{otherwise}\end{cases}$$

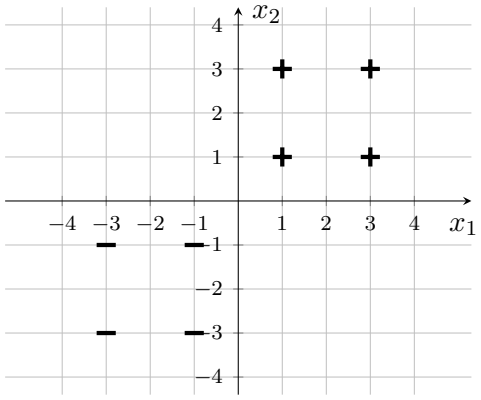
(b) For your selected classifier from part (a), determine the test error rate.

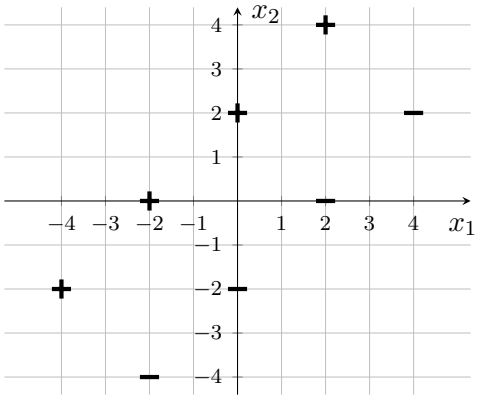
(c) For the two training datasets below, draw the decision boundary for the closest average classifier. Write down the resulting training error for each dataset.





(d) Assume that we use PCA dimensionality reduction to project these training datasets from two dimensions to one dimension, and then apply the closest average classifier to the resulting one-dimensional datasets. Explain how well the closest average classifier will perform for each one-dimensional dataset, and justify your reasoning. You do not need to perform exact calculations, but you should feel free to make sketches to support your reasoning.


x_2
x_1
4
3
2
1
−1
−2
−3
−4
−4
−3
−2
−1
1
2
3
4


x_2
x_1
4
3
2
1
−1
−2
−3
−4
−4
−3
−2
−1
1
2
3
4