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	-

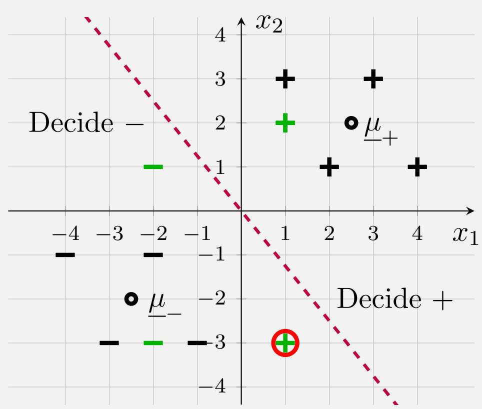
Testing Data

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.

Solution:

The sample mean vectors are $\underline{\mu}_+ = \begin{bmatrix} 2.5 \\ 2 \end{bmatrix}$ and $\underline{\mu}_+ = \begin{bmatrix} -2.5 \\ -2 \end{bmatrix}$, as shown below.



(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.

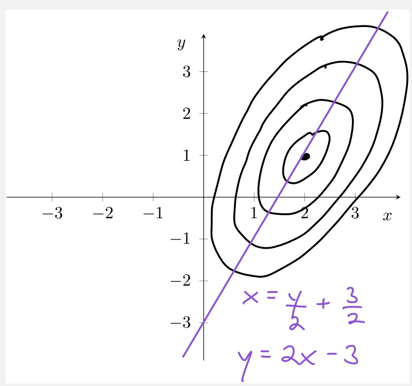
Solution:

There are no training errors. The testing data has been added to the plot above in green. There is one test error, circled in red.

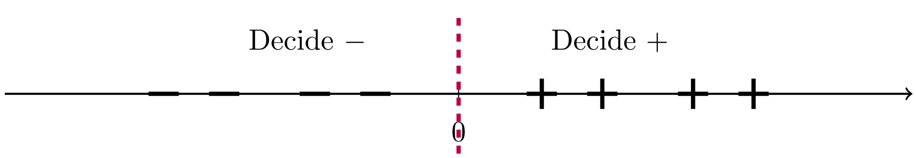
(d) Sketch the Gaussian contour plot for PCA dimensionality reduction applied to the entire training dataset. Draw the eigenvector that would be used to project down from two dimensions to one dimension.

Solution:

The empirical mean vector of the entire training set is equal to $\boldsymbol{\mu} = (0,0)^\top$ since for every training vector $(x_1, x_2)^\top$ in the training set, $(-x_1, -x_2)^\top$ is also a training vector. We can also explicitly and exactly calculate the empirical covariance matrix and it's eigenvectors, but given the symmetric distribution of training set vectors we can qualitatively infer that the direction of greatest variation of values (the first principal direction or the top eigenvector of the empirical covariance matrix) will be approximately aligned with the direction of the $(1,1)^\top$ vector. The exact contour plot and the exact orientation of the principal eigenvector is shown in the figure below. Your qualitative sketch need not match this exactly, but it should be roughly consistent with what is shown below in terms of (i) the elliptical shapes of the contours, (ii) the centers of the counters positioned at the origin, and (iii) the orientations of the major and minor axes of the elliptical contours.

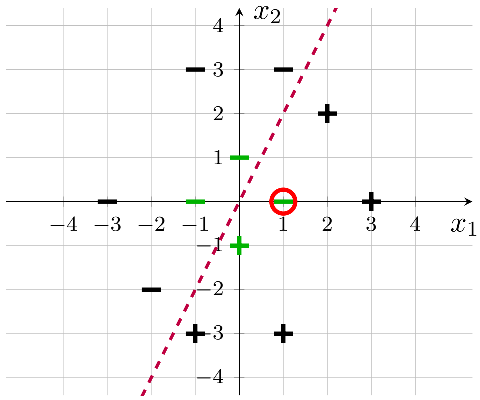


(e) On a one-dimensional axis such as the one below, sketch the pluses and minuses for the training data after PCA dimensionality reduction to a single dimension, using your vector from (d). (You do not need to calculate the exact values for each one-dimensional point.) Sketch the decision boundary for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

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

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

Solution:

$D_4(x_1, x_2)$, which is the only classifier with a training error of zero.

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

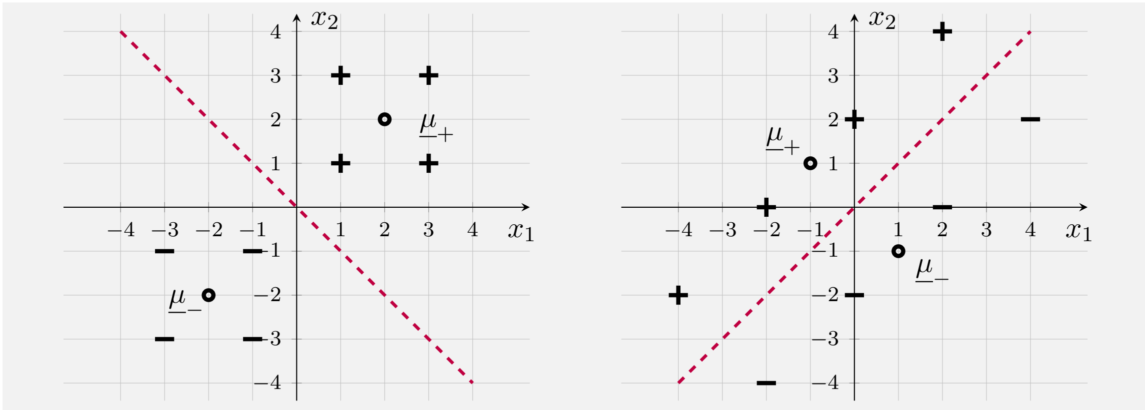
Solution:

The linear classifier boundary for $D_4(x_1, x_2)$ has been added to the plot, along with the testing points in green. Only one point is misclassified, leading to a test error rate of $1/4$.

(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.

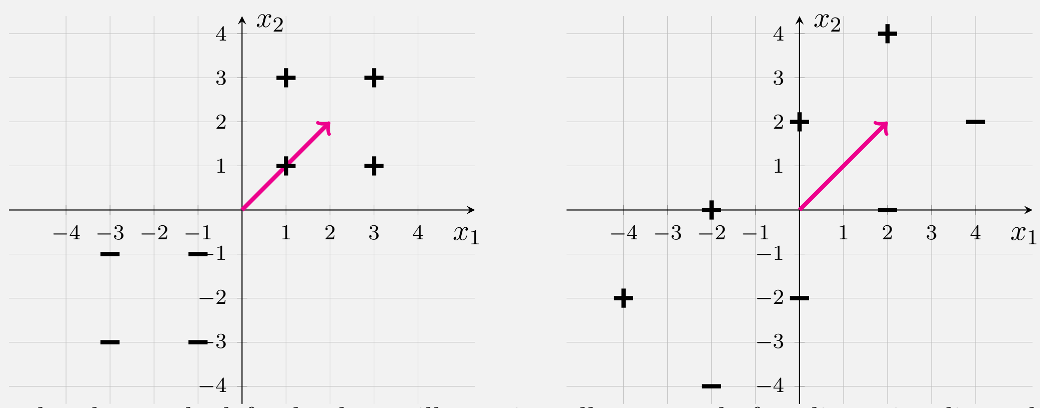
Solution:

For both datasets, the training error rate is zero.

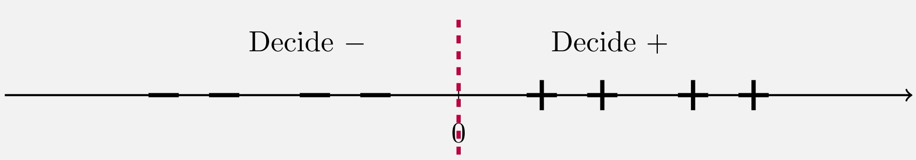


(d) Assume that we use PCA dimensionality reduction to project these training datasets from two dimensions to one dimension, and then apply the closest average classifier to the resulting one-dimensional datasets. Explain how well the closest average classifier will perform for each one-dimensional dataset,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

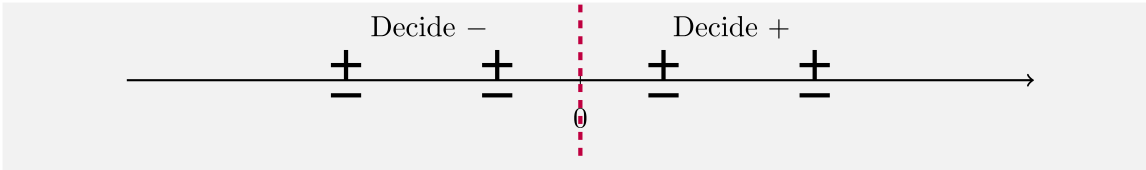
Solution:



For the plot on the left, the data will remain well-separated after dimensionality reduction to one dimension. The PCA vector is shown in purple and the resulting one-dimensional dataset, along with the decision boundary of the closest average classifier is shown below.



For the plot on the right, the data will be poorly separated after dimensionality reduction to one dimension. The PCA vector is shown in purple and the resulting one-dimensional dataset, along with the decision boundary of the closest average classifier is shown below.


Decide −
Decide +
0