Homework 10: finish by 6/22.

Reading: Notes: Chapters 9 - 10 **Videos:** 9.3 - 10.2

Problem 10.1 (**Video 9.1, 9.2, 9.3, 9.4**, **Lecture Problem**)

A biomedical engineer is testing whether their new device improves patient outcomes after knee surgery. They know that the typical patient population with this condition has a mean symptom severity rating of 7.2. As a preliminary study, they examine 100 patients who have received the device and find a reported average sample severity rating of 6.5 with a sample standard deviation of 1.5. Recall that the following useful values

$$\Phi(-1.64) = Q(1.64) = 0.05,\ \Phi(-1.96) = Q(1.96) = 0.025,\ \Phi(-2.57) = Q(2.57) = 0.005\ .$$

(a) Construct the 90% confidence interval for the average symptom severity rating for patients with the new device.

> **Solution:**
>
> Here, $n = 100$, $\mu = 7.2$, $M_{100} = 6.5$, and $V_{100} = 1.5^2 = 2.25$. Since $n > 30$, we can assume M_n is Gaussian and use the Q function instead of the Student's t CDF. We want confidence level $1 - \alpha = 0.9$ so $\alpha/2 = 0.05$. It follows that he 90% confidence interval around the sample mean is
>
> $$M_n \pm \frac{\sqrt{V_n}}{\sqrt{n}} Q^{-1}\left(\frac{\alpha}{2}\right) = 6.5 \pm \frac{\sqrt{2.25}}{\sqrt{100}} \cdot 1.64 = 6.5 \pm 0.246 = [6.254, 6.746]\ .$$

(b) You would like to construct a significance test to determine whether this new device improves patient outcomes. Explain your null hypothesis and select a test from amongst those covered in Video 9.3.

> **Solution:**
>
> The null hypothesis is that the new device does not improve patient outcomes as measured by symptom severity rating. Since we are comparing a sample mean to a known baseline mean, we should use a one-sample test. Since there are more than 30 samples, we can assume a Gaussian distribution, taking the sample variance to be the known variance, and use a one-sample Z-test.

(c) Calculate the appropriate test statistic.

> **Solution:**

We begin by calculating the Z-statistic:

$$Z = \frac{\sqrt{n}(M_n - \mu)}{\sqrt{V_n}} = \frac{\sqrt{100}(6.5 - 7.2)}{1.5} = \frac{10 \cdot -0.7}{\frac{3}{2}} = \frac{-7 \cdot 2}{3} = -\frac{14}{3} = -4.67$$

(d) Based on the observed test data, should the null hypothesis be rejected with significance level 0.05? Explain all steps in your calculation.

Solution:

We can now compare to the value of the Z-statistic that would be required to meet the significance level α exactly. Here, we want $\alpha = 0.05$, which would be attained by any Z with satisfying $|Z| > 1.96$ since $2\Phi(-1.96) = 0.05$. Since we have $Z = -4.67$, this is easily satisfied. Overall, we reject the null hypothesis that the device has the same average symptom severity rating as the population without the device at a significance level of 0.05. Alternatively, you could have compute the $\mathrm{p-value} = 2\Phi(-|Z|) = 2\Phi(-4.67) = 3.0 \times 10^{-6}$ and noted that it is much smaller than the significance level of 0.05.

Problem 10.2 (**Video 9.1, 9.2, 9.3, 9.4**)

The weights in ounces of coffee in bags produced by Rhett's Roastery can be assumed to be i.i.d. from some single underlying Gaussian distribution. The null hypothesis H_0 is that the mean weight of a bag of coffee is 12 ounces.

You purchase 25 bags and weigh them to obtain values $X_1, X_2, \ldots, X_{25}$, from which you calculate sample mean $M_{25} = 11.92$ and sample variance $V_{25} = 0.04$.

Possibly useful:

$$2\Phi(-1.64) = 0.1, \qquad 2\Phi(-1.96) = 0.05, \qquad 2\Phi(-2.57) = 0.01,$$

$$2F_{T_{24}}(-1.71) = 0.1, \qquad 2F_{T_{24}}(-2.06) = 0.05, \qquad 2F_{T_{24}}(-2.80) = 0.01,$$

(a) What type of statistical test should you use to reject or not reject H_0?

Solution:

This is a test of whether the mean is equal to the baseline mean of $\mu = 12$, with variance unknown. (We have merely a *sample* variance.) Thus, we should use a one-sample T-test.

(b) Write an expression for the test statistic to use for the test you selected in part (a).

Solution:

The T-statistic to use is

$$T = \frac{\sqrt{n}(M_n - \mu)}{\sqrt{V_n}} = \frac{\sqrt{25}(11.92 - 12)}{\sqrt{0.04}} = \frac{5(11.92 - 12)}{0.2} = -2.$$

(c) Do you reject or not reject H_0 at significance level 0.05? Do not merely answer yes or no. Instead, write a self-contained, clear sentence that states your null hypothesis, whether you reject it (or fail to reject it), and according to which significance level.

Solution:

We want to check whether the p-value of $2F_{T_{24}}(-2)$ is larger or smaller than 0.05, and we *reject* the null hypothesis when this value is *smaller* than 0.05. From the given values of $2F_{T_{24}}(-2.06) = 0.05$ and $2F_{T_{24}}(-1.71) = 0.1$, we can see that the p-value for our test statistic is larger than 0.05. Overall, we fail to reject the null hypothesis that the mean weight of a bag of coffee is equal to 12 ounces at a significance level of 0.05.

(d) Find a 90% confidence interval for the mean.

Solution:

In the standard notation, seeking a 90% confidence interval means we are using $\alpha = 0.1$. From the data provided, we see that the width of the interval is ± 1.71 standard deviations around the estimated mean value. Since the variance of the one sample is estimated as 0.04, the variance of the sample mean is $\frac{0.04}{25}$ so the standard deviation is 0.04. Thus, the interval is $11.92 \pm 0.04 \times 1.71 = 11.92 \pm 0.0684$.

Alternatively, you can use the formulas in the handout. The confidence interval will be centered at M_n with width 2ε, where

$$\varepsilon = -\frac{\sqrt{V_n} F^{-1}_{T_{n-1}}(\alpha/2)}{\sqrt{n}} = -\frac{\sqrt{0.04} F^{-1}_{T_{24}}(0.05)}{\sqrt{25}} = -\frac{\sqrt{0.04}\,|-1.71|}{\sqrt{25}} = \frac{(0.2)(1.71)}{5} = 0.0684,$$

where we use the first value in the second row of the table. Thus the confidence interval is $[11.92 - 0.0684,\ 11.92 + 0.0684]$.

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

You are given the following 8 training data points and 8 testing data points:

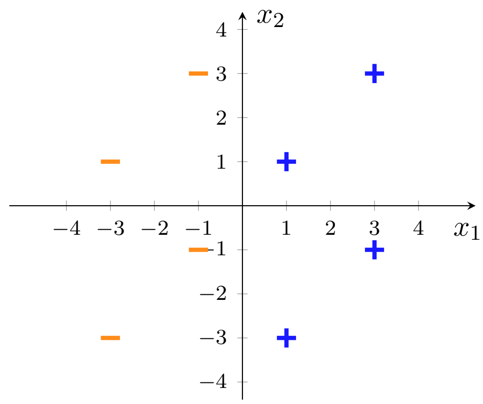
Training Data

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

Testing Data

x_1	x_2	label
2	3	+
0	-1	+
2	1	+
-0.5	-3	+
-2	1	-
-3	-2	-
0.5	3	-
0.5	1	-

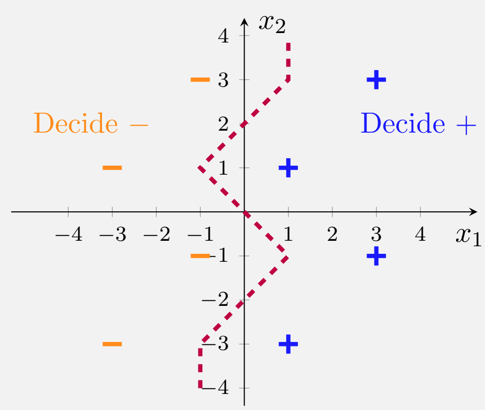
The data is formatted as a table where the first column is the x_1 coordinate (i.e., feature), the second column is the x_2 coordinate (i.e., feature), and the third column is the label, which is either $+1$ or -1. For each part below, you will be asked to create a sketch of the decision region, place the testing points onto the sketch, and make the corresponding decisions. To help get you started, here is a sketch of the training data.



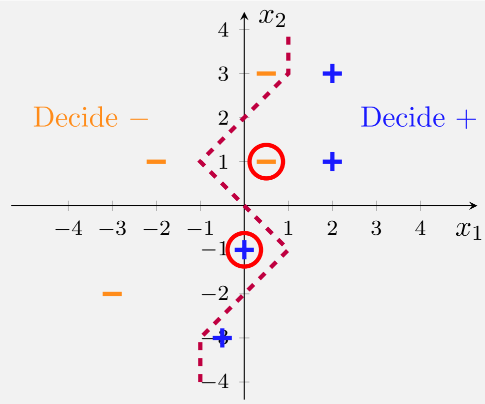
(a) Sketch the decision boundaries for the *nearest neighbor* classifier, add the testing data to your plot, and circle the testing points that will be misclassified. Determine the resulting estimate of the probability of error from the testing data.

Solution:

The nearest neighbor classifier makes its decision by first finding the closest point in the training data to the testing point and then assigning the corresponding label to the testing point. We can visualize this as a decision region by first plotting the training data and finding the regions where we are closest to + training points. Below, we've drawn the resulting boundary as a dashed purple line.



Now, let's keep the decision boundary but plot the testing data. We can see that one + and − will be misclassified and they are circled in red. The probability of error estimate is $2/8 = 1/4$.



(b) Determine the sample mean $\hat{\underline{\mu}}_+$ for the + training data and the sample mean $\hat{\underline{\mu}}_-$ for the − training data. On a single sketch, include the sample means for + and − data, the closest average decision boundary, the testing data, and circle the testing points that will be misclassified. Estimate the probability of error from the testing data.

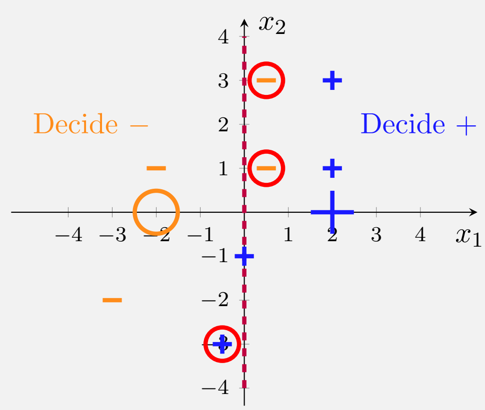
Solution:

The sample mean vectors are

$$\hat{\underline{\mu}}_+ = \frac{1}{4}\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 3 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ -3 \end{bmatrix} + \begin{bmatrix} 3 \\ -1 \end{bmatrix} \right) = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$\hat{\underline{\mu}}_- = \frac{1}{4}\left(\begin{bmatrix} -1 \\ -1 \end{bmatrix} + \begin{bmatrix} -3 \\ 3 \end{bmatrix} + \begin{bmatrix} -1 \\ 3 \end{bmatrix} + \begin{bmatrix} -3 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

We have drawn the sample mean vectors as very large + and − markers on the sketch below. The closest average classifier decides based on whether a testing point is closer to $\hat{\underline{\mu}}_+$ or $\hat{\underline{\mu}}_-$, and the resulting decision boundary is drawn as a dashed purple line. Three testing points will be misclassified and are circled in red. The + exactly on the boundary will not be misclassified since we decide ties in favor of +. The probability of error estimate is 3/8.



(c) Determine the slope a and offset b for the LDA decision boundary, expressed as a line of the form $x_2 = ax_1 + b$. For this training dataset, the pooled estimate of the covariance matrix is

$\hat{\mathbf{\Sigma}} = \frac{4}{3}\begin{bmatrix}1 & 1\\ 1 & 5\end{bmatrix}$ and its inverse is $\hat{\mathbf{\Sigma}}^{-1} = \frac{3}{16}\begin{bmatrix}5 & -1\\ -1 & 1\end{bmatrix}$. On a single sketch, include the sample means for + and − data, the line for the LDA decision boundary, the testing data, and circle the testing points that will be misclassified. Estimate the probability of error from the testing data.

Solution:

The LDA decision rule is

$$D_{\text{LDA}}(\underline{x}) = \begin{cases} +1 & \underline{x}^\mathsf{T}\hat{\mathbf{\Sigma}}^{-1}(\underline{\hat{\mu}}_+ - \underline{\hat{\mu}}_-) \;\geq\; \frac{1}{2}(\underline{\hat{\mu}}_+^\mathsf{T}\hat{\mathbf{\Sigma}}^{-1}\underline{\hat{\mu}}_+ - \underline{\hat{\mu}}_-^\mathsf{T}\hat{\mathbf{\Sigma}}^{-1}\underline{\hat{\mu}}_-), \\ -1 & \text{otherwise.} \end{cases}$$

We begin by plugging in and simplifying:

$$\begin{aligned}
\underline{x}^\mathsf{T}\hat{\mathbf{\Sigma}}^{-1}(\underline{\hat{\mu}}_+ - \underline{\hat{\mu}}_-) &= \begin{bmatrix}x_1 & x_2\end{bmatrix}\frac{3}{16}\begin{bmatrix}5 & -1\\ -1 & 1\end{bmatrix}\left(\begin{bmatrix}2\\0\end{bmatrix} - \begin{bmatrix}-2\\0\end{bmatrix}\right)\\
&= \begin{bmatrix}x_1 & x_2\end{bmatrix}\frac{3}{16}\begin{bmatrix}5 & -1\\ -1 & 1\end{bmatrix}\begin{bmatrix}4\\0\end{bmatrix}\\
&= \begin{bmatrix}x_1 & x_2\end{bmatrix}\frac{3}{4}\begin{bmatrix}5\\-1\end{bmatrix}\\
&= \frac{3}{4}(5x_1 - x_2)
\end{aligned}$$

$$\underline{\hat{\mu}}_+^\mathsf{T}\hat{\mathbf{\Sigma}}^{-1}\underline{\hat{\mu}}_+ - \underline{\hat{\mu}}_-^\mathsf{T}\hat{\mathbf{\Sigma}}^{-1}\underline{\hat{\mu}}_- = \begin{bmatrix}2 & 0\end{bmatrix}\frac{3}{16}\begin{bmatrix}5 & -1\\ -1 & 1\end{bmatrix}\begin{bmatrix}2\\0\end{bmatrix} - \begin{bmatrix}-2 & 0\end{bmatrix}\frac{3}{16}\begin{bmatrix}5 & -1\\ -1 & 1\end{bmatrix}\begin{bmatrix}-2\\0\end{bmatrix} = 0$$

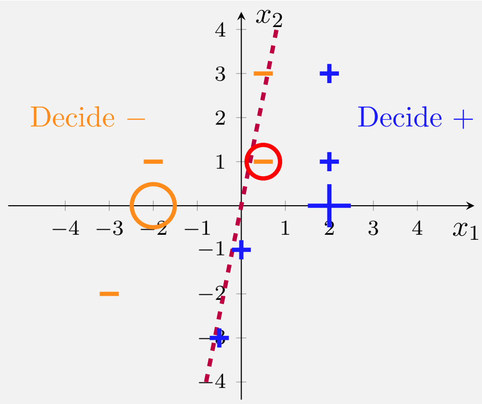
$$\frac{3}{4}(5x_1 - x_2) \geq 0 \implies 5x_1 \geq x_2$$

We end up with the simple decision rule

$$D_{\text{LDA}}(\underline{x}) = \begin{cases} +1 & 5x_1 \geq x_2 \\ -1 & \text{otherwise,} \end{cases}$$

which means the decision boundary line is $x_2 = 5x_2$, i.e., it has slope $a = 5$ and offset $b = 0$.

This decision boundary is sketched below along with the sample mean vectors and the testing data. There is only a single error, circled in red. The probability of error estimate is 1/8.



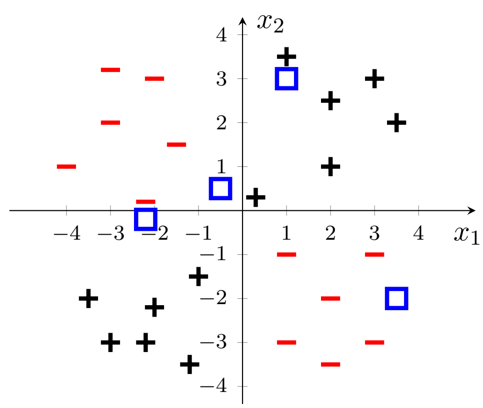
(d) Argue why, for this particular training dataset, the QDA decision boundary will be exactly the same as the LDA decision boundary.

Solution:

Notice that the + training data points minus their sample mean vector $\underline{\hat{\mu}}_+$ are in exactly the same configuration as the − training data points minus their mean vector $\underline{\hat{\mu}}_-$. As a result, the sample covariance matrices $\hat{\mathbf{\Sigma}}_+$ and $\hat{\mathbf{\Sigma}}_-$ will be equal. The LDA decision boundary is the result of an ML rule applied to $\mathcal{N}(\underline{\hat{\mu}}_+, \hat{\mathbf{\Sigma}})$ and $\mathcal{N}(\underline{\hat{\mu}}_- \hat{\mathbf{\Sigma}})$ where $\hat{\mathbf{\Sigma}} = \frac{(n_+-1)\hat{\mathbf{\Sigma}}_+ + (n_- -1)\hat{\mathbf{\Sigma}}_-}{n_+ + n_- - 2} = \hat{\mathbf{\Sigma}}_+ = \hat{\mathbf{\Sigma}}_-$ in this special case. The QDA decision boundary is the result of an ML rule applied to $\mathcal{N}(\underline{\hat{\mu}}_+, \hat{\mathbf{\Sigma}}_+)$ and $\mathcal{N}(\underline{\hat{\mu}}_- \hat{\mathbf{\Sigma}}_-)$. Since $\hat{\mathbf{\Sigma}} = \hat{\mathbf{\Sigma}}_+ = \hat{\mathbf{\Sigma}}_-$, this will yield the same decision boundary as LDA. In other words, LDA presumes the covariance matrices are equal whereas QDA presumes they can be different. If they are actually equal, then QDA and LDA will produce the same boundary.

Problem 10.4 (**Video 10.1, 10.2**)

You are given the 24 training data points on the figure denoted by + and − symbols. You are also given four blue testing points, denoted by squares, for which you have no labels. The coordinates of the blue squares are: Square A = (-2.2, -0.2); Square B = (-0.5, 0.5); Square C = (1 3); Square D = (3.5 -2).



(a) Determine the labels for the four numbered squares that you would obtain using the *nearest neighbor classifier* studied in class.

Solution:

The labels are: Square A = -, Square B = +, Square C = + , Square D = -.

(b) Explain why the *linear discriminant analysis (LDA)* classifier has a high training error rate.

Solution:

Note that the means under the two classes are both very close to the origin. Hence, the distributions will have the same mean and covariance under both hypotheses in LDA, which makes it very hard to *linearly* separate the classes.

(c) Is there a linear classifier that can attain a training error less than 1/8? Justify your answer.

Solution:

No, there isn't one. There is no straight line that will have less than 6 errors (training error $6/24 = 2/8$) in the training data.

(d) Consider a *quadratic* classifier of the form:

$$D_{\mathrm{q}}(x_1, x_2) = \begin{cases} +1, & a \cdot x_1^2 + b \cdot x_1 x_2 + c \cdot x_2^2 \geq 0 \\ -1, & a \cdot x_1^2 + b \cdot x_1 x_2 + c \cdot x_2^2 < 0 \end{cases}$$

where the values $+1$ and -1 correspond to the $+$ and $-$ symbols, respectively. Select values for the coefficients a, b and c that would achieve zero training error. Clearly justify your reasoning. (**Hint:** Don't look for formulas in your sheets. Instead, look at the quadrants that the training points of each class are in.)

Solution:

We can pick $a = c = 0, b = 1$, because all the + points have $x_1 x_2 > 0$, and all the - points have $x_1 x_2 < 0$.

(e) Determine the testing labels for the four numbered squares that you would obtain using the classifier that you specified in part (d).

Solution:

The labels are: Square A = +, Square B = -, Square C = + , Square D = -.