COMS 474
Homework 6

Haadi Majeed

Spring 2022

Problem 1

[8 points]
Recall that for a two-class {0, 1} classification problem with p features $\{X_1, ..., X_p\}$, logistic regression models have the form

$$P(Y = 1|X_1, ..., X_p) = \frac{e^{\beta_0+\beta_1 X_1+...+\beta_p X_p}}{1 + e^{\beta_0+\beta_1 X_1+...+\beta_p X_p}} \qquad P(Y = 0|X_1, ..., X_p) = \frac{1}{1 + e^{\beta_0+\beta_1 X_1+...+\beta_p X_p}}$$

For a model with fitted coefficients $\{\beta_0, \beta_1, ..., \beta_p\}$, what is the boundary between $\hat{Y} = 1$ and $\hat{Y} = 0$? Simply your work as much as possible.

By assigning the equations equal to each other, and since they have identical values in the denominator we can be sitting with $e^{\beta_0+\beta_1 X_1+...+\beta_p X_p} = 1$ and we then assign $X_1, ..., X_P = 0$ along with $\beta_0 = 0$. We can get the following sequence of equations:

$$\frac{e^{\beta_0}}{1 + e^{\beta_0}} = \frac{1}{1 + e^{\beta_0}}$$
$$\frac{1}{1 + 1} = \frac{1}{1 + 1}$$
$$\frac{1}{2} = \frac{1}{2}$$

Resulting with a boundary of $x = \frac{1}{2}$ where the two functions intersect at.

Problem 2

[15 points (8,2,5)]
In this problem you will identify the max margin classifier for the following data set with three samples.

Sample #	Y	X_1	X_2
1	+1	0	2
2	-1	0	0
3	+1	-2	0

You do not need to draw a scatter plot of these samples, though it may help you to check your answers to the following questions.

A

What are the set of coefficients $\{\beta_0, \beta_1, \beta_2\}$ for the line

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0$$

that passes through the two $Y = +1$ samples? Normalize the coefficients so that $\beta_1^2 + \beta_2^2 = 2$ (In lecture we used a different normalization; this is chosen to simplify your work).
Since there are two $Y = +1$ samples, we create a system of equations.

The first sample:

$$\beta_1 + \beta_1(0) + \beta_2(2) = 0$$

$$\beta_0 + 2\beta_2 = 0$$

The second sample:

$$\beta_1 + \beta_1(-2) + \beta_2(0) = 0$$

$$\beta_0 - 2\beta_1 = 0$$

With this, we can determine the following:

β_0	β_1	β_2
2	1	-1

This satisfies the normalization equation resulting with:

$$-1^2 + 1^2 = 2$$

B

Confirm whether your coefficients $\{\beta_0, \beta_1, \beta_2\}$ of $\beta_0 + \beta_1 X_1 + \beta_2 X_2$ evaluates positive or negative for the $Y = -1$ sample.

If the evaluation is positive, change your formula so that it evaluates as negative for that sample.

The coefficients

β_0	β_1	β_2
2	1	-1

We can enter into the formula:
$2 + 1 * 0 + -1 * 0 = 2$
Since it is positive, we can modify all β values while still maintaining validity for part A by doing the following:
$-1 * \{\beta_0, \beta_1, \beta_2\}$

C

Using your results from A and B, guess what the optimal separating line (i.e. the boundary line for the max margin classifier)

$$\widetilde{\beta_0} + \widetilde{\beta_1} X_1 + \widetilde{\beta_2} X_2 = 0$$

is with $\widetilde{\beta_1^2} + \widetilde{\beta_2^2} = 2$.
Confirm that using the classifier

$$\hat{Y}(X_1, X_2) = sign\left(\widetilde{\beta_0} + \widetilde{\beta_1} X_1 + \widetilde{\beta_2} X_2\right)$$

achieves zero classification error and that all training samples have the same magnitude of $\widetilde{\beta_0} + \widetilde{\beta_1} X_1 + \widetilde{\beta_2} X_2$.

Using properties from A and B We can fill in the formula

$$\hat{Y}(X_1, X_2) = -\left(\widetilde{2} + \widetilde{1} X_1 - \widetilde{1} X_2\right)$$

When we plug in the values for each of the samples, we get the following:
Sample 1: $-(2 + 1 * 0 - 1 * 2 = 0)$
Sample 2: $-(2 + 1 * 0 - 1 * 0 = -2)$
Sample 3: $-(2 + 1 * -2 - 1 * 0 = 0)$
Thus we need to adjust the value of β_0 from 2 to 1 in order to get a more equal distribution.
Sample 1: $-(1 + 1 * 0 - 1 * 2 = 1)$
Sample 2: $-(1 + 1 * 0 - 1 * 0 = -1)$
Sample 3: $-(1 + 1 * -2 - 1 * 0 = 1)$
This modified version now has a correct sign of negative one, and confirms the new classifier with equal magnitudes and zero error.

Problem 3

[24 points (15, 4, 2, 3)]
You are working on a two-class {A,B} classification problem using two features $\{X_1, X_2\}$. You decide to use QDA. You estimate the mean vectors for each class and find that they are the same,

$$\widehat{\mu_A} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \widehat{\mu_B} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

When you estimate the covariance matrices, you find they are both diagonal matrices,

$$\widehat{\sum_A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \widehat{\sum_B} = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$$

Lastly, the number of samples for Class A and B are the same, so $\widehat{\pi_A} = \frac{1}{2} = \widehat{\pi_B}$.

A

What is the equation for the QDA decision boundary for this problem? Simplify it as much as possible.

B

Sketch a plot by hand showing the boundary. Indicate for what region(s) $\hat{Y} = A$ and for what region(s) $\hat{Y} = B$

C

What would the prediction be for the mean value of both conditional distributions, $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$?
What would the prediction be for the point $\begin{bmatrix} 3 \\ -3 \end{bmatrix}$?

D

Suppose we had a third class C, also centered at the origin and with a diagonal covariance matrix,

$$\widehat{\mu_C} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \widehat{\sum_C} = \begin{bmatrix} 9 & 0 \\ 0 & 9 \end{bmatrix}$$

and that all classes had equal priors, $\pi_A = \pi_B = \pi_C = \frac{1}{3}$.
Sketch a plot similar to B, of the decision regions for this problem with three classes (e.g. make it clear on the plot what region has $\hat{Y} = A$, what region has $\hat{Y} = B$, and what region has $\hat{Y} = C$). You do not need to do calculations, but extrapolate from what you just observed for the two class problem.

Fig 4.28

$$\delta_K = -\frac{1}{2} x^T \Sigma_K^{-1} x + x^T \Sigma_K^{-1} \mu_K - \frac{1}{2} \mu_K^T \Sigma_K^{-1} \mu_K - \frac{1}{2} \log|\Sigma_K| + \log \pi_K$$

$$-\frac{1}{2}[x_1\ x_2]\begin{bmatrix}1&0\\0&1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} + [x_1\ x_2]\begin{bmatrix}1&0\\0&1\end{bmatrix}\begin{bmatrix}0\\0\end{bmatrix} - \frac{1}{2}[0\ 0]\begin{bmatrix}1&0\\0&1\end{bmatrix}\begin{bmatrix}0\\0\end{bmatrix} - \frac{1}{2}\log\begin{vmatrix}1&0\\0&1\end{vmatrix} + \log\frac{1}{2}$$

$$-\frac{1}{2}[x_1^2\ \ x_2^2] - \frac{1}{2}\log\begin{bmatrix}1&0\\0&1\end{bmatrix} + \log\frac{1}{2} = \delta_A$$

$$-\frac{1}{2}[4x_1^2\ \ 4x_2^2] - \frac{1}{2}\log\begin{bmatrix}4&0\\0&4\end{bmatrix} + \log\frac{1}{2} = \delta_B$$

Fig 4.25

$$\delta_A = \delta_B$$

$$-\frac{1}{2}[x_1^2\ \ x_2^2] - \frac{1}{2}\log\begin{bmatrix}1&0\\0&1\end{bmatrix} + \log\frac{1}{2} = -\frac{1}{2}[4x_1^2\ 4x_2^2] - \frac{1}{2}\log\begin{bmatrix}4&0\\0&4\end{bmatrix} + \log\frac{1}{2}$$

$$[x_1^2\ x_2^2] - \frac{1}{2}\log\begin{bmatrix}1&0\\0&1\end{bmatrix} = [4x_1^2\ \ 4x_2^2]\ \log\begin{bmatrix}4&0\\0&4\end{bmatrix}$$

$$[-3x_1^2\ \ -3x_2^2] = \log\begin{bmatrix}3&0\\0&3\end{bmatrix}$$

I was not able to make much progress past this point, I just ended up getting answers that seemed more and more wrong :| .