CS 4980/6980: Introduction to Data Science © Spring 2018

Lecture 11: Solution for Mid-Exam

INSTRUCTOR: DANIEL L. PIMENTEL-ALARCÓN Scribed by: Shannon Lee & Xumin Cai

This is preliminary work and has not been reviewed by instructor. If you have comments about typos, errors, notation inconsistencies, etc., please email the scribers.

11.1 Norm

Problem 1.

Consider:

$$\mathbf{x}=\begin{bmatrix}8\\-3\\0\\1\\-7\end{bmatrix}, \qquad \mathbf{y}=\begin{bmatrix}-5\\9\\-4\\6\\-2\end{bmatrix}$$

Compute

$$\|\mathbf{x}-\mathbf{y}\|_1 \quad \text{and} \quad \|\mathbf{x}-\mathbf{y}\|_2$$

Solution:

$$\begin{aligned}\|\mathbf{x}-\mathbf{y}\|_1 &= |(8+5)|+|(-3-9)|+|(0+4)|+|(1-6)|+|(-7+2)| \\ &= |13|+|-12|+|4|+|-5|+|-5| \\ &= 39\end{aligned}$$

$$\begin{aligned}\|\mathbf{x}-\mathbf{y}\|_2 &= \sqrt{(8+5)^2+(-3-9)^2+(0+4)^2+(1-6)^2+(-7+2)^2} \\ &= \sqrt{13^2+-12^2+4^2+-5^2+-5^2} \\ &= \sqrt{379}\end{aligned}$$

11.2 Linear Regression

Problem 2.

Consider the following feature matrix containing information about two features (height and weight) of three individuals:

$$\mathbf{X} = \begin{bmatrix} 180 & 150 & 170 \\ 165 & 175 & 165 \end{bmatrix}$$

Suppose we also have the following information of a variable of interest (glucose level) of the same individuals:

$$\mathbf{Y} = \begin{bmatrix} 110 & 140 & 180 \end{bmatrix}$$

Find $\beta \in \mathbb{R}^2$ that minimizes $\|(\mathbf{Y} - \beta^T \mathbf{X})^T\|_2$.

Solution:

We use $\hat{\beta} = (\mathbf{X}\mathbf{X}^T)^{-1}(\mathbf{X}\mathbf{Y}^T)$ to solve for β. We have...

$$\mathbf{X} = \begin{bmatrix} 180 & 150 & 170 \\ 165 & 175 & 165 \end{bmatrix} \quad and \quad \mathbf{X}^T = \begin{bmatrix} 180 & 165 \\ 150 & 175 \\ 170 & 165 \end{bmatrix}$$

$$\mathbf{X}\mathbf{X}^T = \begin{bmatrix} 180 & 150 & 170 \\ 165 & 175 & 165 \end{bmatrix} \begin{bmatrix} 180 & 165 \\ 150 & 175 \\ 170 & 165 \end{bmatrix} = \begin{bmatrix} 83800 & 81300 \\ 81300 & 80350 \end{bmatrix}$$

$$(\mathbf{X}\mathbf{X}^T)^{-1} = \begin{bmatrix} 6.499x10^{-4} & -6.576x10^{-4} \\ -6.576x10^{-4} & 6.778x10^{-4} \end{bmatrix}$$

Further... we have

$$\mathbf{Y}^T = \begin{bmatrix} 110 \\ 140 \\ 180 \end{bmatrix}$$

Thus,

$$\mathbf{X}\mathbf{Y}^T = \begin{bmatrix} 180 & 150 & 170 \\ 165 & 175 & 165 \end{bmatrix} \begin{bmatrix} 110 \\ 140 \\ 180 \end{bmatrix} = \begin{bmatrix} 71400 \\ 70700 \end{bmatrix}$$

Lastly,

$$(\mathbf{X}\mathbf{X}^T)^{-1}(\mathbf{X}\mathbf{Y}^T) = \begin{bmatrix} 6.499x10^{-4} & -6.576x10^{-4} \\ -6.576x10^{-4} & 6.778x10^{-4} \end{bmatrix} \begin{bmatrix} 71400 \\ 70700 \end{bmatrix} = \begin{bmatrix} -.088 \\ .969 \end{bmatrix}$$

Therefore,

$$\hat{\beta} = (\mathbf{X}\mathbf{X}^T)^{-1}(\mathbf{X}\mathbf{Y}^T) = \begin{bmatrix} -.088 \\ .969 \end{bmatrix}$$

11.3 Logistic Regression

Problem 3.

Recall that in logistic regression we use gradient descent to find a vector βthat classifies a data point $\mathbf{x}$ according to:

$$\underbrace{\frac{1}{1+e^{-\beta^T \mathbf{x}}}}_{P(y=1|\mathbf{x})} \quad \overset{\hat{y}=1}{\underset{\hat{y}=0}{\gtrless}} \quad \underbrace{1 - \frac{1}{1+e^{-\beta^T \mathbf{x}}}}_{P(y=0|\mathbf{x})}$$

Suppose

$$\beta = \begin{bmatrix} 8 \\ -2 \\ -7 \\ 4 \\ 9 \\ 0 \\ -6 \end{bmatrix}$$

According to this β, how would the following points be classified?

$$\mathbf{x}_1 = \begin{bmatrix} 8 \\ -2 \\ -7 \\ 4 \\ 9 \\ 0 \\ -6 \end{bmatrix} \quad \mathbf{x}_2 = \begin{bmatrix} -8 \\ 2 \\ 7 \\ -4 \\ -9 \\ 0 \\ 6 \end{bmatrix} \quad \mathbf{x}_3 = \begin{bmatrix} -1 \\ 0 \\ 8 \\ -5 \\ 3 \\ 4 \\ -3 \end{bmatrix} \quad \mathbf{x}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ -1 \\ 9 \\ -2 \end{bmatrix}$$

Solution:

Step One, caculate $\beta^T \mathbf{x}_1$:

$$\begin{aligned}
\beta^T \mathbf{x}_1 &= \begin{bmatrix} 8 & \text{-2} & \text{-7} & 4 & 9 & 0 & \text{-6} \end{bmatrix} \begin{bmatrix} 8 \\ \text{-2} \\ \text{-7} \\ 4 \\ 9 \\ 0 \\ \text{-6} \end{bmatrix} \\
&= 64 + 24 + 49 + 81 + 0 + 36 \\
&= 250
\end{aligned}$$

We can tell that $\underbrace{\frac{1}{1+e^{-250}}}_{P(y=1|\mathbf{x}_1)} = 1$ and $\underbrace{1 - \frac{1}{1+e^{-250}}}_{P(y=0|\mathbf{x}_1)} = 0$. So $\mathbf{x}_1$ is labeled as $\{1\}$.

Step Two caculate $\beta^T \mathbf{x}_2$:

$$\begin{aligned}
\beta^T \mathbf{x}_2 &= \begin{bmatrix} 8 & \text{-2} & \text{-7} & 4 & 9 & 0 & \text{-6} \end{bmatrix} \begin{bmatrix} \text{-8} \\ 2 \\ 7 \\ \text{-4} \\ \text{-9} \\ 0 \\ 6 \end{bmatrix} \\
&= -64 - 24 - 49 - 81 + 0 - 36 \\
&= -250
\end{aligned}$$

We can tell that $\underbrace{\frac{1}{1+e^{250}}}_{P(y=1|\mathbf{x}_2)} = 0$ and $\underbrace{1 - \frac{1}{1+e^{250}}}_{P(y=0|\mathbf{x}_2)} =1$. So $\mathbf{x}_2$ is labeled as $\{0\}$.

Step Three, caculate $\beta^T \mathbf{x}_3$:

$$\begin{aligned}\beta^T \mathbf{x}_3 &= \begin{bmatrix} 8 & \text{-2} & \text{-7} & 4 & 9 & 0 & \text{-6} \end{bmatrix} \begin{bmatrix} \text{-1} \\ 0 \\ 8 \\ \text{-5} \\ 3 \\ 4 \\ \text{-3} \end{bmatrix} \\ &= -8 + 0 - 56 - 20 + 27 + 0 + 18 \\ &= -39 \end{aligned}$$

We can tell that $\underbrace{\frac{1}{1+e^{39}}}_{P(y=1|\mathbf{x}_3)} = 0$ and $\underbrace{1 - \frac{1}{1+e^{39}}}_{P(y=0|\mathbf{x}_3)} = 1$. So $\mathbf{x}_3$ is labeled as $\{0\}$.

Step Four, caculate $\beta^T \mathbf{x}_4$:

$$\begin{aligned}\beta^T \mathbf{x}_4 &= \begin{bmatrix} 8 & \text{-2} & \text{-7} & 4 & 9 & 0 & \text{-6} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ \text{-1} \\ 9 \\ \text{-2} \end{bmatrix} \\ &= 0 + 0 + 0 + 4 - 9 + 0 + 12 \\ &= 7 \end{aligned}$$

We can tell that $\underbrace{\frac{1}{1+e^{-7}}}_{P(y=1|\mathbf{x}_4)} = 1$ and $\underbrace{1 - \frac{1}{1+e^{-7}}}_{P(y=0|\mathbf{x}_4)} = 0$. So $\mathbf{x}_4$ is labeled as $\{1\}$.

11.4 Nearest Neighbour

Problem 4.

Consider the following data matrix

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 \end{bmatrix} = \begin{bmatrix} 4 & 3 & \text{-7} & 1 \\ \text{-2} & \text{-1} & 4 & 3 \\ 3 & 1 & \text{-6} & 5 \\ \text{-1} & \text{-4} & 2 & 1 \end{bmatrix}$$

whose columns correspond to the following classes $\{1, 1, 2, 3\}$, respectively. According to the nearest neighbor algorithm, how would the following points be classified?

$$\mathbf{Y} = [\mathbf{y}_1 \quad \mathbf{y}_2] = \begin{bmatrix} 4 & \text{-1} \\ 2 & 4 \\ 3 & \text{-5} \\ 1 & 3 \end{bmatrix}$$

Solution:

Step One, find minimum distance between $\mathbf{y}_1$ and each $\mathbf{x}_i$

$$\begin{aligned} \|\mathbf{y}_1 - \mathbf{x}_1\|_2 &= \sqrt{(4-4)^2 + (2+2)^2 + (3-3)^2 + (1+1)^2} \\ &= \sqrt{0 + 16 + 0 + 4} \\ &= \sqrt{20} \end{aligned}$$

$$\begin{aligned} \|\mathbf{y}_1 - \mathbf{x}_2\|_2 &= \sqrt{(4-3)^2 + (2+1)^2 + (3-1)^2 + (1+4)^2} \\ &= \sqrt{1 + 9 + 4 + 25} \\ &= \sqrt{39} \end{aligned}$$

$$\begin{aligned} \|\mathbf{y}_1 - \mathbf{x}_3\|_2 &= \sqrt{(4+7)^2 + (2-4)^2 + (3+6)^2 + (1-2)^2} \\ &= \sqrt{121 + 4 + 81 + 1} \\ &= \sqrt{207} \end{aligned}$$

$$\begin{aligned} \|\mathbf{y}_1 - \mathbf{x}_4\|_2 &= \sqrt{(4-1)^2 + (2-3)^2 + (3-5)^2 + (1-1)^2} \\ &= \sqrt{9 + 1 + 4 + 0} \\ &= \sqrt{14} \end{aligned}$$

We could find that the minimum value is $\|\mathbf{y}_1 - \mathbf{x}_4\|_2$ equals $\sqrt{14}$. So the class of $\mathbf{y}_1$ is $\{3\}$.

Step Two, find minimum distance between $\mathbf{y}_2$ and each $\mathbf{x}_i$

$$\begin{aligned}
\|\mathbf{y}_2 - \mathbf{x}_1\|_2 &= \sqrt{(-1-4)^2 + (4+2)^2 + (-5-3)^2 + (3+1)^2} \\
&= \sqrt{25 + 36 + 64 + 16} \\
&= \sqrt{141}
\end{aligned}$$

$$\begin{aligned}
\|\mathbf{y}_2 - \mathbf{x}_2\|_2 &= \sqrt{(-1-3)^2 + (4+1)^2 + (-5-1)^2 + (3+4)^2} \\
&= \sqrt{16 + 25 + 36 + 49} \\
&= \sqrt{126}
\end{aligned}$$

$$\begin{aligned}
\|\mathbf{y}_2 - \mathbf{x}_3\|_2 &= \sqrt{(-1+7)^2 + (4-4)^2 + (-5+6)^2 + (3-2)^2} \\
&= \sqrt{36 + 0 + 1 + 1} \\
&= \sqrt{38}
\end{aligned}$$

$$\begin{aligned}
\|\mathbf{y}_2 - \mathbf{x}_4\|_2 &= \sqrt{(-1-1)^2 + (4-3)^2 + (-5-5)^2 + (3-1)^2} \\
&= \sqrt{4 + 1 + 100 + 4} \\
&= \sqrt{109}
\end{aligned}$$

We could find that the minimum value is $\|\mathbf{y}_2 - \mathbf{x}_3\|_2$ equals $\sqrt{38}$. So the class of $\mathbf{y}_2$ is $\{2\}$.