DS 100/200: Principles and Techniques of Data Science **Date: March 13, 2020**

Discussion #8

Name: Junior

Linear Regression Fundamentals

1. In this problem, we will review some of the core concepts in linear regression.

   (a) Suppose we create a linear model with parameters $\hat{\theta} = [\hat{\theta}_0, \ldots, \hat{\theta}_p]$. As we saw in lecture, given an observation $\boldsymbol{x}$, such a model makes predictions $\hat{y} = \hat{\theta} \cdot \boldsymbol{x}$.
   Suppose $\hat{\theta} = [2, 0, 1]$ and we receive an observation $\boldsymbol{x_1} = [1, 2, 3]$. What $\hat{y}_1$ value will this model predict for the given observation?

   $$\hat{\theta} \cdot x = [2,0,1] \cdot [1,2,3]$$
   $$= 2(1) + 0(2) + (1)(3) = 2+3 = \boxed{5}$$

   (b) Suppose the correct y_1 was 3.5. What will be the L_2 loss for our prediction $\hat{y}_1$ from question 1a?

   L_2: $(\text{actual} - \text{pred})^2$

   $$= (3.5 - 5)^2 = 1.5^2 = \boxed{2.25}$$

   (c) Suppose we receive another observation $\boldsymbol{x_2} = [-2, 5, 1]$. What $\hat{y}_2$ value will this model predict for the given observation?

   $$[2,0,1] \cdot [-2,5,1] = 2(-2) + 0(5) + (1)(1)$$
   $$= -4+1 = \boxed{-3}$$

   (d) Suppose the correct y_2 for 1c was was -4. What will be the mean squared error of the $\hat{\theta}$ from 1a given the two observations (from 1b and 1c)?

   $$MSE(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2}{2}$$
   $$= \frac{2.25 + (-4-(-3))^2}{2}$$
   $$= \frac{2.25+1}{2} = \frac{3.25}{2} = \boxed{1.625}$$

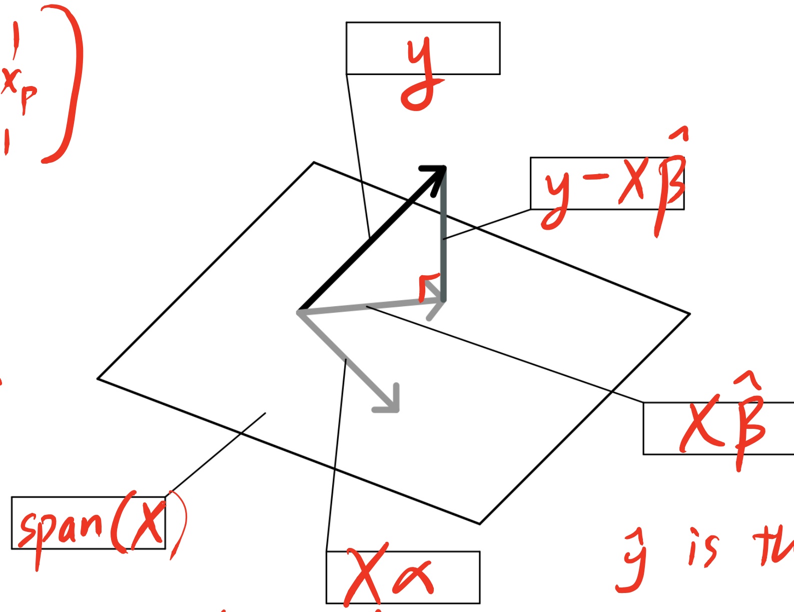
Geometry of Least Squares

$\beta = \theta$

2. Suppose we have a dataset represented with the design matrix span($\mathbb{X}$) and response vector $\vec{y}$. We use linear regression to solve for this and obtain optimal weights as $\hat{\beta}$. Draw the geometric interpretation of the column space of the design matrix span($\mathbb{X}$), the response vector $\vec{y}$, the residuals $\vec{y} - \mathbb{X}\hat{\beta}$, and the predictions $\mathbb{X}\hat{\beta}$.

$$X = \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_p \\ | & | & & | \end{bmatrix}$$

Xv : linear combination of vectors in span(X)



$\hat{y} = X\hat{\beta}$

$\hat{y}$ is the projection of y onto span(X)

α is any vector that's not $\hat{\beta}$

(a) What is always true about the residuals in least squares regression? Select all that apply.

☑ A. They are orthogonal to the column space of the design matrix. $\rightarrow X$

☑ B. They represent the errors of the predictions.

☐ C. Their sum is equal to the mean squared error.

☐ D. Their sum is equal to zero.

☐ E. None of the above.

$r = y - \hat{y}$

$MSE = \frac{1}{n}\sum (y_i - \hat{y}_i)^2$

$r \neq MSE$

(b) Which are true about the predictions made by OLS? Select all that apply.

☑ A. They are projections of the observations onto the column space of the design matrix.

☑ B. They are linear in the features.

☑ C. They are orthogonal to the residuals.

☐ D. They are orthogonal to the column space of the features.

☐ E. None of the above.

OLS : ordinary least squares :

$\hat{y} = X\theta$

$L = \frac{1}{n}\|y - X\theta\|_2^2$

Modeling

1. We wish to model exam grades for DS100 students. We collect various information about student habits, such as how many hours they studied, how many hours they slept before the exam, and how many lectures they attended and observe how well they did on the exam. Propose a model to predict exam grades and a loss function to measure the performance of your model on a single student.

x_1 : hours studied

x_2 : hours slept

x_3 : lectures attended

$$\hat{y} = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_0$$

$$L = \frac{1}{n} \Sigma (y_i - \hat{y}_i)^2$$

2. Suppose we collected even more information about each student, such as their eye color, height, and favorite food. Do you think adding these variables as features would improve our model?

$\boxed{\text{NO}}$