MALIS
Group Exercise
October 11 2022

Group Name:	
Group Members:	Solutions

Linear Regression and Maximum Likelihood Estimation

1. Complete the proof from the slide deck on MLE, slide 25:

$$\frac{\partial}{\partial \sigma^2}\left(\frac{N}{2}\log\sigma^2 + \frac{1}{2\sigma^2}(y-Xw)^T(y-Xw)\right) = 0$$

$$\frac{N}{\cancel{2}}\cdot\frac{1}{\sigma^2} - \frac{1}{\cancel{2}(\sigma^2)^2}(y-Xw)^T(y-Xw) = 0$$

$$\frac{N}{\sigma^2} = \frac{1}{(\sigma^2)^2}(y-Xw)^T(y-Xw)$$

$$\frac{(\sigma^2)^{\cancel{2}}}{\cancel{\sigma^2}} = \frac{1}{N}(y-Xw)^T(y-Xw) \Rightarrow \sigma^2 = \frac{1}{N}(y-Xw)^T(y-Xw)$$

2. Suppose we have a data set with five feature, x_1 = GPA, x_2 =IQ, x_3 = Level (1 for College and 0 for High School), x_4 = Interaction between GPA and IQ, and x_5 = Interaction between GPA and Level. The output variable is starting salary after graduation (in thousands of dollars). Suppose we fit the model, and get $\hat{w}_0 = 50$, $\hat{w}_1 = 20$, $\hat{w}_2 = 0.07$, $\hat{w}_3 = 35$, $\hat{w}_4 = 0.01$, $\hat{w}_5 = -10$.

   a. Write down the expression of the linear model

Answering using GPA, IQ, ... is also correct

$$\hat{y} = h(x) = 50 + 20x_1 + 0.07x_2 + 35x_3 + 0.01\underbrace{x_1x_2}_{x_4} - 10\underbrace{x_1x_3}_{x_5}$$

   b. Which answer is correct, and why?
      i. For a fixed value of IQ and GPA, high school graduates earn more, on average, than college graduates.
      ii. For a fixed value of IQ and GPA, college graduates earn more, on average, than high school graduates.
      $\boxed{\text{iii.}}$ For a fixed value of IQ and GPA, high school graduates earn more, on average, than college graduates provided that the GPA is high enough.
      iv. For a fixed value of IQ and GPA, college graduates earn more, on average, than high school graduates provided that the GPA is high enough.

- Do $\hat{y}_{HS}(x) > \hat{y}_G$
- You will get an expression that allows to obtain the GPA score needed for high school graduates to earn more.
- It is true as long as GPA > 3.5

c. Predict the salary of a college graduate with IQ of 110 and a GPA of 4.0.

$$\hat{y} = 50 + 20(4) + 0.07(110) + 35(1) + 0.01(110 * 4) - 10 * 4 * 1$$

$$= 137.1 \text{ K USD}$$

3. I collect a set of data (n = 100 observations) containing a single input feature and a quantitative response. I then fit a linear regression model to the data, as well as a separate cubic regression, i.e.

$$y = w_0 + w_1x + w_2x^2 + w_3x^3 + \epsilon.$$

a. Suppose that the true relationship between X and Y is linear, i.e. $y = w_0 + w_1x + \epsilon$. Consider the training residual sum of squares (RSS) for the linear regression, and also the training RSS for the cubic regression. Would we expect one to be lower than the other, would we expect them to be the same, or is there not enough information to tell? Justify your answer.

Polynomial terms will allow a bett fit of the noise in the training data so, it will have a lower RSS.

b. Answer (a) using test rather than training RSS.

As the true relationship is linear, the linear model ~~should~~ $\hat{y} = \hat{w}_0 + \hat{w}_1 x$ ~~we~~ should have a lower RSS in the test set.

Credits: Excercises 2 and 3 have been adapted from Introduction to Statistical Learning, Ch 3.