ESE 541 Week 3 and 4 Practice Problems - Solutions

Problem 1. Multiply the following matrices.

1. $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$

Row 1, column 1 is given by $1 \times 1 + 0 \times 3 = 1$. Row 1, column 2 is given by $1 \times 2 + 0 \times 4 = 2$. Row 2, column 1 is given by $0 \times 1 + 1 \times 3 = 3$. Row 2, column 2 is given by $0 \times 2 + 1 \times 4 = 4$. The matrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is known as the identity matrix. Thus, $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$.

2. $\begin{bmatrix} 9 & 6 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 9 & 6 \\ 3 & 2 \end{bmatrix}$

$\begin{bmatrix} 9 & 6 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 9 & 6 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 99 & 66 \\ 33 & 22 \end{bmatrix}$. This is called the "lucky fool" matrix. Sadly, this property does not hold for all matrices. Challenge: Using Python, try to find all possible 2x2 matrix pairs that exhibit this property.

3. $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^3$

$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^3 = \begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix}$. Try out different powers of this matrix. In general, $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^n = \begin{bmatrix} F_{n+1} & F_n \\ F_n & F_{n-1} \end{bmatrix}$. This matrix is called the Fibonacci matrix and has exquisite mathematical beauty.

Problem 2. Using the equations below for $\hat{\beta_1}$ and $\hat{\beta_0}$, argue that in the case of simple linear regression, the least squares line always pass through the point ($\bar{x}$, $\bar{y}$).

$$\hat{\beta_1} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\hat{\beta_0} = \bar{y} - \hat{\beta_1}\bar{x}$$

where $\bar{x}$ and $\bar{y}$ are the empirical means.

The least squares regression can be written as $\hat{Y} = f(X) = \hat{\beta_0} + \hat{\beta_1}X$. Thus, if $\bar{x}$ is our input, we have $\hat{\beta_1} \to 0$ and $\hat{\beta_0} \to \bar{y}$ and we have $f(\bar{x}) = \bar{y}$.

Problem 3. For the following exercise, consider the following linear model:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

where ϵ is a Gaussian random variable following a normal distribution N(0,σ^2). Using this model, we obtain the following training dataset:

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{3} = \{(-1, -2), (0, 0), (1, 5)\}.$$

1. Find the coefficients for the model, $\hat{\beta_0}$ and $\hat{\beta_1}$.

The empirical mean is $\bar{x} = 0$ and $\bar{y} = 1$. Thus, we have $\hat{\beta_1} = \frac{(-1-0)(-2-1)+(0-0)(0-1)+(1-0)(5-1)}{(-1-0)^2+(0-0)^2+(1-0)^2} = 3.5$ and $\hat{\beta_0} = 1 - 3.5 \times 0 = 1$. Thus, our linear equation is $\bar{Y} = 3.5X + 1$.

2. Compute the RSS of the linear model.

Using the empirical results, our linear model yields $\{(-1, -2.5), (0, 1), (1, 4.5)\}$. The RSS is thus $(-2 - (-2.5))^2 + (0 - 1)^2 + (5 - 4.5)^2 = 1.5$ using the equation

$$RSS = \sum_{i=1}^{3} (y_i - \hat{y}_i)^2$$

3. Compute the RSE of the linear model.

Since we have one predictor, we have

$$RSE = \sqrt{\frac{RSS}{N - p - 1}} = \sqrt{\frac{1.5}{3 - 1 - 1}} = \sqrt{1.5}$$

4. Compute the TSS of the linear model.

$$TSS = \sum_{i=1}^{3} (y_i - \bar{y}) = (-2 - 1)^2 + (0 - 1)^2 + (5 - 1)^2 = 26$$

5. Compute the R^2 of the linear model.

$$R^2 = \frac{TSS - RSS}{TSS} = \frac{26 - 1.5}{26} = 0.94$$

Problem 4. Consider a dataset with $n = 100$ points containing a single predictor ($p = 1$) and a quantitative response Y. To this dataset, we fit a linear regression model $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$, as well as a separate cubic regression, i.e. $\hat{Y}' = \hat{\beta}'_0 + \hat{\beta}'_1 X + \hat{\beta}'_2 X^2 + \hat{\beta}'_3 X^3$.

1. Assume that nature is using a linear model to generate the dataset. Would you expect the training RSS of the linear regression to be smaller or larger than the training RSS of the cubic regression? Justify your answer.

We would expect the training RSS of the linear regression to be larger than that of the cubic regression, because the cubic regression is more flexible.

2. Answer (**1**) using test RSS instead of training RSS.

We would expect the testing RSS of the linear regression to be smaller than that of the cubic regression, because the true distribution is linear. A model that is too flexible may overfit the data.

3. Assume that nature is not using a linear model to generate the data, but we don't know how far it is from linear. Consider the training 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.

We would expect the training RSS of the cubic regression to be lower than that of the linear regression regardless of the true distribution, because the cubic regression is more flexible.

4. Answer (**3**) using test RSS instead of training RSS.

There is not enough information to determine which testing RSS would be lower, because we are not certain of the true distribution.

Problem 5. Assume that 'nature' behaves according to the following linear additive model: $Y = \beta_0 + \beta_1 X + \varepsilon$. After collecting a dataset $\mathcal{D}$ from nature, we would like to analyze the following hypothesis:

$$H_1\colon \beta_1 = 1 \text{ and } H_2\colon \beta_1 = 2.$$

Note: $SE\left(\hat{\beta}_1\right)$ is the same as $SD\left(\hat{\beta}_1\right)$.

1. Assuming that you compute $\hat{\beta}_1 = -3$ and $SE\left(\hat{\beta}_1\right) = 1$, what conclusions can you draw about the above hypothesis?

$\beta_1 \in [-3 \pm 2(1)]$, so $\beta_1 \in [-5, -1]$. None of the hypotheses are in this range, so we reject $H_1$and reject H_2 at a significance level of $\alpha = 0.05$.

2. Assuming that you compute $\hat{\beta}_1 = 1/2$ and $SE\left(\hat{\beta}_1\right) = 2$, what conclusions can you draw about the above hypothesis?

$\beta_1 \in [1/2 \pm 2(2)]$, so $\beta_1 \in [-7/2, 9/2]$. Both of the hypotheses are in this range, so we fail to reject $H_1$and fail to reject H_2 at a significance level of $\alpha = 0.05$.

3. Assuming that you compute $\hat{\beta}_1 = 2$ and $SE\left(\hat{\beta}_1\right) = 1/2$, what conclusions can you draw about the above hypothesis?

$\beta_1 \in [2 \pm 2(1/2)]$, so $\beta_1 \in [1, 3]$. Both of the hypotheses are in this range, so we fail to reject $H_1$and fail to reject H_2 at a significance level of $\alpha = 0.05$.

Problem 6. We would like to learn the parameters of a simple quadratic predictor of the form $\hat{Y} = \alpha X^2 + \gamma$. In particular, find the value of α that minimizes the RSS given the following dataset:

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^3 = \left\{(0,0), \left(1, \frac{1}{2}\right), \left(-1, \frac{3}{2}\right)\right\}.$$

Define the variable $Z = X^2$. Then $\mathcal{D} = \{(z_i, y_i)\}_{i=1}^3 = \left\{\left(0^2, 0\right), \left(1^2, \frac{1}{2}\right), \left((-1)^2, \frac{3}{2}\right)\right\}$ and $\hat{Y} = \alpha Z + \gamma$. The coefficient α and intercept γ that minimize the RSS are found using least squares regression:

$$\alpha = \frac{\sum_{i=1}^n (z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^n (z_i - \bar{z})^2}$$

$$\gamma = \bar{y} - \alpha \bar{z}$$

Since $\bar{z} = \frac{0+1+1}{3} = \frac{2}{3}$ and $\bar{y} = \frac{0+1/2+3/2}{3} = \frac{2}{3}$, we find that $\alpha = 1$.