

Homework 1

Sean Eva

January 2022

1 Part I - Theoretical Problems

1. We are not able to perform the QR factorization in this scenario. Since the $\text{rank}(A) < p$ we know that the column vectors of A are not all linearly independent because in order for the column vectors to all be linearly independent we need that $\text{rank}(A) = p$. QR factorization has the requirement that the column vectors of A to be linearly independent.
2. In order to solve this we use $A^T A x = A^T b$. Then we get that

$$\begin{aligned} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}^T \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} x &= \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}^T \begin{bmatrix} 0 \\ 0 \\ 6 \end{bmatrix} \\ \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} x &= \begin{bmatrix} 6 \\ 18 \end{bmatrix} \\ \begin{bmatrix} \frac{7}{3} & -1 \\ -1 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} x &= \begin{bmatrix} \frac{7}{3} & -1 \\ -1 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 6 \\ 18 \end{bmatrix} \\ x^* &= \begin{bmatrix} -4 \\ 3 \end{bmatrix}. \end{aligned}$$

By the definition of the least squares solution to the linear system, we know that $Ax^* = \text{proj}_V(b)$ where $V = \text{span}(v_1, \dots, v_n)$ such that v_1, \dots, v_n are the column vectors of A .

$$3. \quad (a) \quad A_n = \begin{bmatrix} 1 & \sin(2\pi/n) & \cos(2\pi/n) \\ 1 & \sin(4\pi/n) & \cos(4\pi/n) \\ \dots & \dots & \dots \\ 1 & \sin(2\pi) & \cos(2\pi) \end{bmatrix}.$$

$$\text{Then } A_n \begin{bmatrix} c_n \\ p_n \\ q_n \end{bmatrix} = y = \begin{bmatrix} 1 & \sin(2\pi/n) & \cos(2\pi/n) \\ 1 & \sin(4\pi/n) & \cos(4\pi/n) \\ \dots & \dots & \dots \\ 1 & \sin(2\pi) & \cos(2\pi) \end{bmatrix} \begin{bmatrix} c_n \\ p_n \\ q_n \end{bmatrix}$$

(b) I will first calculate

$$\begin{aligned}
A^T A &= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \sin(2\pi/n) & \sin(4\pi/n) & \cdots & \sin 2\pi \\ \cos(2\pi/n) & \cos(4\pi/n) & \cdots & \cos 2\pi \end{bmatrix} \begin{bmatrix} 1 & \sin(2\pi/n) & \cos(2\pi/n) \\ 1 & \sin(4\pi/n) & \cos(4\pi/n) \\ \cdots & \cdots & \cdots \\ 1 & \sin(2\pi) & \cos(2\pi) \end{bmatrix} = \\
&= \begin{bmatrix} n & \sum_{i=1}^n \sin(i2\pi/n) & \sum_{i=1}^n \cos(i2\pi/n) \\ \sum_{i=1}^n \sin(i2\pi/n) & \sum_{i=1}^n \sin^2(i2\pi/n) & \sum_{i=1}^n \sin(i2\pi/n) \cos(i2\pi/n) \\ \sum_{i=1}^n \cos(i2\pi/n) & \sum_{i=1}^n \sin(i2\pi/n) \cos(i2\pi/n) & \sum_{i=1}^n \cos^2(i2\pi/n) \end{bmatrix} = \\
&= \begin{bmatrix} n & 0 & 0 \\ 0 & n/2 & 0 \\ 0 & 0 & n/2 \end{bmatrix}. \text{ Therefore, } \lim_{n \rightarrow \infty} \frac{2\pi}{n} A_n^T A_n = \lim_{n \rightarrow \infty} \frac{2\pi}{n} \begin{bmatrix} n & 0 & 0 \\ 0 & n/2 & 0 \\ 0 & 0 & n/2 \end{bmatrix} = \\
&= \begin{bmatrix} 2\pi & 0 & 0 \\ 0 & \pi & 0 \\ 0 & 0 & \pi \end{bmatrix}. \text{ Then if we recognize that the least squares solu-} \\
&\text{tion states that } A^A x = A^T y \text{ then we know that } \lim_{n \rightarrow \infty} \frac{2\pi}{n} A^T y = \\
&\lim_{n \rightarrow \infty} \frac{2\pi}{n} A^T A x = \begin{bmatrix} 2\pi & 0 & 0 \\ 0 & \pi & 0 \\ 0 & 0 & \pi \end{bmatrix} \begin{bmatrix} c_n \\ p_n \\ q_n \end{bmatrix} = \begin{bmatrix} 2\pi c_n & 0 & 0 \\ 0 & \pi p_n & 0 \\ 0 & 0 & \pi q_n \end{bmatrix}
\end{aligned}$$

(c) I am confused on what to do here.

(d) I am confused on what to do here.

4. (a) Obs 1: $= \sqrt{(0-0)^2 + (3-0)^2 + (0-0)^2} = 3$
Obs 2: $= \sqrt{(2-0)^2 + (0-0)^2 + (0-0)^2} = 2$
Obs 3: $= \sqrt{(0-0)^2 + (1-0)^2 + (3-0)^2} = \sqrt{10}$
Obs 4: $= \sqrt{(0-0)^2 + (1-0)^2 + (2-0)^2} = \sqrt{5}$
Obs 5: $= \sqrt{(-1-0)^2 + (0-0)^2 + (1-0)^2} = \sqrt{2}$
Obs 6: $= \sqrt{(1-0)^2 + (1-0)^2 + (1-0)^2} = \sqrt{3}$

(b) We notice that the nearest neighbor to the test point $(0,0,0)$ is Obs 5 with Euclidean distance $\sqrt{2}$. Since Obs 5 is Green we predict that the test point will also be Green.

(c) The three nearest neighbors to the test point $(0,0,0)$ are Obs 5, 6, and 2 with distanced $\sqrt{2}$, $\sqrt{3}$, and 2 respectively. Obs 5 is Green but Obs 6 and Obs 2 are both Red which means the test point is most likely Red.

(d) A highly nonlinear Bayes Boundary would suggest that there is a less advantages to generalizing further due to higher variance. Therefore, the best value for K in our nearest neighbor analysis would be small.

5. 3. (a) (iii) is the only true one because if the GPA is around 4.0 then the offset from $\hat{\beta}_5$ will decrease the amount more than the gain from $\hat{\beta}_3$ from the college level. Therefore (i) is wrong because it isn't always true for all other values. (ii) is wrong because the split isn't even between the college and high school graduates. Lastly, (iv) is wrong because the higher GPA actually makes the college students earn less than the high school students.

- (b) The salary of a college graduate with IQ of 110 and a GPA of 4.0 is, $\hat{y} = 50 + 20(4) + 0.07(110) + 35(1) + 0.01(4*110) + -10(4*1) = 137.1$
 - (c) False: It is also dependent on the standard error of the beta estimates.
4. (a) Without knowing more details about the training data, it is difficult to know which training RSS is lower between linear and cubic. However, as the true relationship between x and y is linear we may expect the least squares line to be closer to the true regression line and consequently, the RSS for the linear regression may be lower than for the cubic regression.
- (b) In this case, the test RSS depends upon the test data. Therefore, we have not enough information to form a conclusion. However, we may assume that polynomial regression will have a higher test RSS as the overfit from training would have more error than the linear regression.
- (c) Polynomial regression has the lower train than RSS, then the linear fit because of higher flexibility. No matter what the underlying true relationship is the more flexible model will more closely follow points and reduce train RSS.
- (d) There is not enough information to tell which test RSS would be lower for either regression given the problem statement is such that we don't know how far it is from being linear. If it is closer to linear than to cubic, the linear regression test RSS could be lower than the cubic regression test RSS. Alternatively, if it is closer to cubic than linear, then the cubic regression test RSS could be lower than the linear regression test RSS. It is due to bias resistance trade off, it is not clear what level of flexibility will fit the data better.
5. Given that $y_i = x_i\beta + e_i; i = 1, \dots, n$. Then the least squares estimate of β is obtained by minimizing $\Delta = \sum_{i=1}^n (y_i - x_i\beta)^2$.

$$\begin{aligned}\frac{\partial \Delta}{\partial \beta} &= - \sum_{i=1}^n (y_i - x_i\beta)x_i \\ &= 0 \\ \hat{\beta} &= \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}.\end{aligned}$$

Now,

$$\begin{aligned}
 \hat{y}_i &= x_i \hat{\beta} \\
 &= x_i x \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} \\
 &= \frac{\sum_{i=1}^n x_i x_i y_i}{\sum_{i=1}^n x_i^2} \\
 &= \sum_{i=1}^n \frac{x_i x_i}{\sum_{i=1}^n x_i^2} x y_i \\
 &= \sum_{i=1}^n a_i y_i,
 \end{aligned}$$

where $a_i = \frac{x_i x_i}{\sum_{i=1}^n x_i^2}; i = 1, 2, \dots, n$.

6. The least square line equation is $y = \hat{\beta}_0 + \hat{\beta}_1 x$, so if we substitute in \bar{x} for x we obtain that, $y = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x} = \bar{y}$. Therefore, we can conclude that the least squares line passes through the point (\bar{x}, \bar{y}) .

2 Part II - Programming

1. 1-3. In hw1.py file
 4. The regression line fits the data seemingly well. However it appears from the data that we may want to consider using a logistic model to make a better fitting regression line.
2.
 1. We can say that the pairs mpg and displacement, mpg and weight, cylinder and displacement, cylinders and weight, and displacement and weight are highly correlated.
 2. I am confused on what to do here.
 3. I am confused on what to do here.
- 3a)-(c) In hw1.py file
 - (d) They are linearly related to each other with a large collection of points in the middle which is expected since they are collected according to a normal distribution.
 - (e) Both $\hat{\beta}_0, \hat{\beta}_1$ are very close to both β_0, β_1 since the data output is very close to linear.
 - (f) In hw1.py file
 - (g)-(i) I am confused on what to do here.