STATS 790 Assignment 2

Yiran Zhang
400119421

10 February, 2023

Question 1

(a) In linear regression, the assumed model is $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \epsilon$, it can be written in matrix form as $\overrightarrow{Y} = X\overrightarrow{\beta} + \overrightarrow{\epsilon}$, where $\overrightarrow{\beta}$ is the vector of coefficients (dimension is (p+1) $\times$ 1), Y is the n $\times$ 1 vector of responses, X is an n $\times$ (p+1) matrix of predictors, it is also assumed to be full rank. The estimator of $\overrightarrow{Y}$ is $\hat{\overrightarrow{Y}} = X\overrightarrow{\beta}$.

Naive Linear Algebra Use the least square estimation:

$$
\begin{aligned}
RSS(\overrightarrow{\beta}) &= \sum_{i=1}^{n} \epsilon_i^2 \\
&= \sum_{i=1}^{n} (y_i - \hat{y_i})^2 \\
&= (\overrightarrow{Y} - \hat{\overrightarrow{Y}})^T (\overrightarrow{Y} - \hat{\overrightarrow{Y}}) \\
&= (\overrightarrow{Y} - X\overrightarrow{\beta})^T (\overrightarrow{Y} - X\overrightarrow{\beta})
\end{aligned}
$$

We want to minimize the residual sum of square value, therefore we take the first derivative of it and set it to 0.

$$
\begin{aligned}
\frac{\partial RSS(\overrightarrow{\beta})}{\partial \overrightarrow{\beta}} &= \frac{\partial (\overrightarrow{Y} - X\overrightarrow{\beta})^T (\overrightarrow{Y} - X\overrightarrow{\beta})}{\partial \overrightarrow{\beta}} \\
&= -2X^T\overrightarrow{Y} + 2X^T X\overrightarrow{\beta} \\
&= 0
\end{aligned}
$$

Therefore, $X^T\overrightarrow{Y} = X^T X\overrightarrow{\beta}$, solve for $\overrightarrow{\beta}$ and get that the coefficients are $\overrightarrow{\beta} = (X^T X)^{-1} X^T \overrightarrow{Y}$. We can also compute the second partial derivative to check if it is minimum: $\frac{\partial^2 RSS(\overrightarrow{\beta})}{\partial \overrightarrow{\beta}^2} = 2X^T X > 0$. Hence the least square is minimized when $\overrightarrow{\beta} = (X^T X)^{-1} X^T \overrightarrow{Y}$.

QR Decomposition Based on the definition of QR Decomposition, since X is defined as full rank, it can be written as the form X=QR, Q is an n × (p+1) orthogonal matrix (i.e., $Q^T = Q^{-1}$), and R is an (p+1) × (p+1) invertible upper triangular matrix. Then we can substitute X in the above equation with Q and R.

$$
\begin{aligned}
\overrightarrow{\beta} &= (X^T X)^{-1} X^T \overrightarrow{Y} \\
&= ((QR)^T QR)^{-1} (QR)^T R^{-1} Q^T \overrightarrow{Y} \\
&= (R^T Q^T QR)^{-1} R^T Q^T \overrightarrow{Y} \\
&= (R^T R)^{-1} R^T Q^T \overrightarrow{Y} \\
&= R^{-1} (R^T)^{-1} R^T Q^T \overrightarrow{Y} \\
&= R^{-1} Q^T \overrightarrow{Y}
\end{aligned}
$$

The coefficients are $\overrightarrow{\beta} = R^{-1} Q^T \overrightarrow{Y}$.

SVD X is full rank, therefore we can apply singular value decomposition to X by writing it in the form $X = U \Sigma V^T$, where U is n × n orthogonal matrix ($U^T U = I$), Σ is n × (p+1) matrix, V is (p+1) × (p+1) orthogonal matrix ($V^T V = I$).

Then, solve for $X \overrightarrow{\beta} = \overrightarrow{Y}$

$$
\begin{aligned}
U \Sigma V^T \overrightarrow{\beta} &= \overrightarrow{Y} \\
U^T U \Sigma V^T \overrightarrow{\beta} &= U^T \overrightarrow{Y} \\
\Sigma V^T \overrightarrow{\beta} &= U^T \overrightarrow{Y} \\
V^T \overrightarrow{\beta} &= \Sigma^{-1} U^T \overrightarrow{Y} \\
V V^T \overrightarrow{\beta} &= V \Sigma^{-1} U^T \overrightarrow{Y} \\
\overrightarrow{\beta} &= V \Sigma^{-1} U^T \overrightarrow{Y}
\end{aligned}
$$

The coefficients are $\overrightarrow{\beta} = V \Sigma^{-1} U^T \overrightarrow{Y}$, note that Σ^{-1} is a pseudo inverse of Σ.

Cholesky Decomposition Cholesky decomposition states that a positive definite matrix can be written as the product of a lower triangular matrix and its transpose. X is full rank, hence $X^T X$ is positive definite, then it can be written as $X^T X = LL^T$ where L is lower triangular.

$$
\begin{aligned}
X\overrightarrow{\beta} &= \overrightarrow{Y} \\
X^T X\overrightarrow{\beta} &= X^T\overrightarrow{Y} \\
LL^T\overrightarrow{\beta} &= X^T\overrightarrow{Y} \\
\overrightarrow{\beta} &= (LL^T)^{-1}X^T\overrightarrow{Y}
\end{aligned}
$$

The coefficients are $\overrightarrow{\beta} = (LL^T)^{-1}X^T\overrightarrow{Y}$, when solving for $\overrightarrow{\beta}$, we can solve for $L^T\overrightarrow{\beta}$ first from the equation $L(L^T\overrightarrow{\beta}) = X^T\overrightarrow{Y}$, then solve for $\overrightarrow{\beta}$ from $L^T\overrightarrow{\beta}$, since using triangular matrix in calculation will be simpler than calculating an inverse.

(b)

(c)

Question 2

Question 3

ESL 3.6

ESL 3.19

ESL 3.28

ESL 3.30

Reference