

STATS 790 Assignment 2

Yiran Zhang
400119421

10 February, 2023

Question 1

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

Naive Linear Algebra Use the least square estimation:

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

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

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

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

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 $n \times (k+1)$ orthogonal matrix (i.e., $Q^T = Q^{-1}$), and R is an $(k+1) \times (k+1)$ invertible upper triangular matrix. Then we can substitute X in the above equation with Q and R .

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

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

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 $n \times n$ matrix, Σ is $n \times (k+1)$ matrix, V is $(k+1) \times (k+1)$ matrix.

Cholesky Decomposition

(b)

(c)

Question 2

Question 3

ESL 3.6

ESL 3.19

ESL 3.28

ESL 3.30

Reference