Homework 4

Aiden Kenny
STAT GR5205: Linear Regression Models
Columbia University

November 9, 2020

Question 1

Question 2

Question 3

Question 4

Collaborators: None

(a) For this question we will assume the matrix $\mathbf{X}$ is *centered*, so that each predictor has mean zero. In n-dimensional Euclidean vector space, for any $\mathbf{v} \in \mathbb{R}^n$ we have $\|\mathbf{v}\|^2 = \mathbf{v}^T\mathbf{v}$, so the ridge regression loss function becomes

$$Q(\boldsymbol{\beta}; \lambda) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|^2 = \mathbf{y}^T\mathbf{y} - 2\boldsymbol{\beta}\mathbf{X}^T\mathbf{y} + \boldsymbol{\beta}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} + \lambda\boldsymbol{\beta}^T\boldsymbol{\beta}.$$

Differentiating the loss function with respect to $\boldsymbol{\beta}$ gives us

$$\frac{\partial Q}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} + 2\lambda\boldsymbol{\beta} = -2\mathbf{X}^T\mathbf{y} + 2\big(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}\big)\boldsymbol{\beta} \overset{\text{set}}{=} \mathbf{0},$$

and finally solving for $\boldsymbol{\beta}$ gives us $\hat{\boldsymbol{\beta}}_\lambda = \big(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}\big)^{-1}\mathbf{X}^T\mathbf{y}$.

(b) When $\lambda = 0$ we have $\hat{\boldsymbol{\beta}}_{\lambda=0} = \big(\mathbf{X}^T\mathbf{X}\big)^{-1}\mathbf{X}^T\mathbf{y}$, the OLS estimator.