

Homework 1 (Written)

Problem 1

For the multiple regression model:

$$f(\mathbf{x}) = \beta_0 + \sum_{j=1}^n \beta_j x_j$$

where x_1, \dots, x_n are inputs and β_0, \dots, β_n are coefficients of the model. Suppose the training set include N samples: $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)$, where $\mathbf{x}_i = (x_{i1}, \dots, x_{in})^T$.

(a) Show that the residual sum of squares

$$\text{RSS}(\boldsymbol{\beta}) = \sum_{i=1}^N \left(y_i - \beta_0 - \sum_{j=1}^n \beta_j x_{ij} \right)^2 \quad (1)$$

can be written in matrix form as

$$\text{RSS}(\boldsymbol{\beta}) = (\mathbf{y} - X\boldsymbol{\beta})^T (\mathbf{y} - X\boldsymbol{\beta})$$

where $\mathbf{y} = (y_1, y_2, \dots, y_N)^T$, $\boldsymbol{\beta} = (\beta_0, \dots, \beta_n)^T$, and X is the matrix:

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1n} \\ 1 & x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{N1} & x_{N2} & \cdots & x_{Nn} \end{bmatrix}$$

(b) Show that the partial derivative of RSS with respect to $\boldsymbol{\beta}$ is:

$$\frac{\partial \text{RSS}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2X^T \mathbf{y} + 2X^T X \boldsymbol{\beta}$$

Use the following matrix calculus results:

$$\frac{\partial \mathbf{x}^T A \mathbf{x}}{\partial \mathbf{x}} = 2A\mathbf{x}$$

if A is symmetric.

(c) Show that the least squares solution (normal equation):

$$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \mathbf{y}$$

minimizes RSS.

Problem 2

Consider using Ridge Regression for modeling. Use the following form of cost function (equivalent to Eqn. 5):

$$J(\boldsymbol{\beta}) = \sum_{i=1}^N \left(y_i - \beta_0 - \sum_{j=1}^n \beta_j x_{ij} \right)^2 + \alpha \sum_{i=1}^n \beta_i^2$$

where all the notations from Problem 1 are still valid here.

- (a) Show that $\sum_{i=1}^n \beta_i^2$ can be written in matrix form as:

$$\sum_{i=1}^n \beta_i^2 = \boldsymbol{\beta}^T A \boldsymbol{\beta}$$

where A is the $(n+1) \times (n+1)$ identity matrix except with a 0 in the top-left cell.

- (b) Show that the closed-form solution for the Ridge Regression is given by

$$\hat{\boldsymbol{\beta}} = (X^T X + \alpha A)^{-1} X^T \mathbf{y}$$

Problem 3 (MATH 5027 Only)

Consider the Ridge Regression problem where the cost function is given in Problem 2. Show that the problem is equivalent to the problem

$$\hat{\boldsymbol{\beta}}^c = \arg \min_{\boldsymbol{\beta}^c} \left\{ \sum_{i=1}^N \left(y_i - \beta_0^c - \sum_{j=1}^n (x_{ij} - \bar{x}_j) \beta_j^c \right)^2 + \alpha \sum_{j=1}^n (\beta_j^c)^2 \right\}$$

where \bar{x}_j denotes the mean of the j th feature of all samples:

$$\bar{x}_j = \frac{1}{N} \sum_{i=1}^N x_{ij}$$

Give the correspondence between $\boldsymbol{\beta}^c$ and the original $\boldsymbol{\beta}$ in the cost function $J(\boldsymbol{\beta})$ in Problem 2.