

STATISTICAL AND MACHINE LEARNING
ECON576
Problem Set 2

1. (2 points) Consider a special linear regression model in which each observation has its own slope coefficient:

$$y_i = x_i\beta_i + \varepsilon_i, \quad x_i \neq 0, \quad i = 1, \dots, n.$$

What is the least squares estimator of β_4 ?

2. (5 points) Consider the following linear regression model with only one covariate:

$$y_i = x_i\beta + \varepsilon_i, \quad i = 1, \dots, n.$$

One popular shrinkage estimator of β is the ridge estimator, defined as:

$$\hat{\beta}_{\text{ridge}} = \underset{\beta}{\operatorname{argmin}} \left[\sum_{i=1}^n (y_i - x_i\beta)^2 + \lambda\beta^2 \right], \quad (1)$$

where $\lambda > 0$ is a penalty term. That is, $\hat{\beta}_{\text{ridge}}$ penalizes any deviation from 0.

- (a) Find the derivative of the objective function with respect to β .
 - (b) Derive an explicit formula for $\hat{\beta}_{\text{ridge}}$.
 - (c) What happens if $\lambda \rightarrow 0$? What happens if $\lambda \rightarrow \infty$?
3. (8 points) This exercise uses gradient descent to solve the ridge minimization problem in (1). Specifically, let $f(\beta)$ denote the objective function in (1), i.e.,

$$f(\beta) = \sum_{i=1}^n (y_i - x_i\beta)^2 + \lambda\beta^2.$$

Next, let $\nabla f(\beta)$ denote the gradient of f , i.e., $\nabla f(\beta) = df/d\beta$. Then, given a positive learning rate $\alpha > 0$ and an initial value β_0 , the gradient descent iteratively updates β_{t+1} using:

$$\beta_{t+1} = \beta_t - \alpha \nabla f(\beta) \Big|_{\beta=\beta_t}.$$

The iteration stops until the successive values of β_{t+1} remain about the same, say, when $|\beta_{t+1} - \beta_t| < \varepsilon$ for some small tolerance level $\varepsilon > 0$.

Now suppose we have the data

$$\mathbf{X} = \begin{pmatrix} 2 \\ 4 \\ 6 \\ -2 \\ -4 \\ -6 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 2.5 \\ 5.8 \\ 3.7 \\ -1.1 \\ -3.7 \\ -5 \end{pmatrix}.$$

Use gradient descent to find the ridge estimate with $\lambda = 10$. Use the initial value $\beta_0 = 0$, $\alpha = 0.001$ and $\varepsilon = 10^{-4}$. What is the ridge estimate? Plot the values of the intermediate β_t .

4. (5 points) This question involves the `Auto` dataset, which is included in the package `ISLR2`.
- (a) Fit a multiple linear regression with `mpg` as the response and all other variables except `name` as predictors. Use the `summary()` to print the results.
 - (b) Is there a relationship between the predictors and the response?
 - (c) For which of the predictors can you reject the null hypothesis $H_0 : \beta_j = 0$?
 - (d) If I obtain a p-value of 0.03 for the null hypothesis $H_0 : \beta_j = 0$, is it correct to interpret it as: given our sample, $\beta_j = 0$ with probability 0.03? Explain.