# Problem Set #2

Ian Bach

2024-08-30

Question #1 1).(2 points) Consider a special linear regression model in which each observation has its own slope coefficient: $y_i = x_i\beta_i + \epsilon_i$, $x_i \neq 0$, $i = 1, \dots, n$. What is the least squares estimator of $\beta_4$?

- $y_i = x_i\beta_i + \epsilon_i$
- $y_i$ is the dependent variable
- $x_i$ is the independent variable
- $\beta_i$ is the slope coefficient specific to the $i$-th observation
- $\epsilon_i$ is the error term

To estimate $\beta_4$, focus on the fourth observation: - $y_4 = x_4\beta_4 + \epsilon_4$

The least squares method seeks to minimize the sum of squared residuals. The residual for the 4th observation:

- $\text{Residual}_4 = y_4 - x_4\beta_4$

The sum of squared residuals:

- $\text{RSS}(\beta_4) = (y_4 - x_4\beta_4)^2$

To find the minimum, we take the derivative of $\text{RSS}(\beta_4)$ with respect to $\beta_4$, set it to zero, and simplify:

- $y_4 = x_4\beta_4$

The least squares estimate $\hat{\beta}_4 = \frac{y_4}{x_4}$

Question #2

(A) Find the derivative of the objective function with respect to $\beta$.

The derivative of a function gives us the slope, or how the function changes as $\beta$ changes. To minimize the function, we need to find where the slope is zero.

Sum of Squared Errors:

- $\frac{\partial}{\partial \beta}\left(\sum_{i=1}^{n} (y_i - x_i\beta)^2\right)$

- The derivative of $(y_i - x_i\beta)^2$ with respect to $\beta$ involves taking the derivative of the inner term $(y_i - x_i\beta)$ and then multiplying by the derivative of the square:
- $-2\sum_{i=1}^{n}(y_i - x_i\beta)$

Penalty:

The penalty term $\lambda\beta^2$ penalizes large values of $\beta$.

- $\frac{\partial}{\partial\beta}(\lambda\beta^2) = 2\lambda\beta$

Combining the Derivatives:

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

(B) Derive an explicit formula for $\hat{\beta}_{\text{ridge}}$.

Setting the Derivative to Zero:

- $\frac{\partial f(\beta)}{\partial\beta} = -2\sum_{i=1}^{n} x_i(y_i - x_i\beta) + 2\lambda\beta = 0$

Simplify:

Divide the equation by 2:

- $-\sum_{i=1}^{n} x_i(y_i - x_i\beta) + \lambda\beta = 0$

Rearrange the equation to collect all terms involving $\beta$ on one side:

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

Factor out $\beta$ on the right side:

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

Solve for $\beta$:

- $\hat{\beta}\text{ridge} = \frac{\sum i=1^n x_i y_i}{\sum_{i=1}^{n} x_i^2 + \lambda}$

(C) What happens if $\lambda \to 0$? What Happens if $\lambda \to \infty$?

- When $\lambda$ approaches zero, the penalty term $\lambda\beta^2$ becomes negligible. This means that the Ridge Regression estimator approaches the Ordinary Least Squares (OLS) estimator.
- When $\lambda \to \infty$, the Ridge Regression estimator approaches zero, meaning the model completely shrinks the coefficient and effectively ignores the input.

Question #3

Consider the following problem related to Ridge Regression.

Define the objective function for Ridge Regression:

Given:

The objective function to minimize is $f(\beta) = \sum_{i=1}^{n}(y_i - x_i\beta)^2 + \lambda\beta^2$,

where: $y_i$ is the dependent variable $x_i$ is the independent variable $\beta$ is the coefficient to be estimated $\lambda$ is the regularization parameter

Define the gradient of the function: The gradient of the objective function is given by:

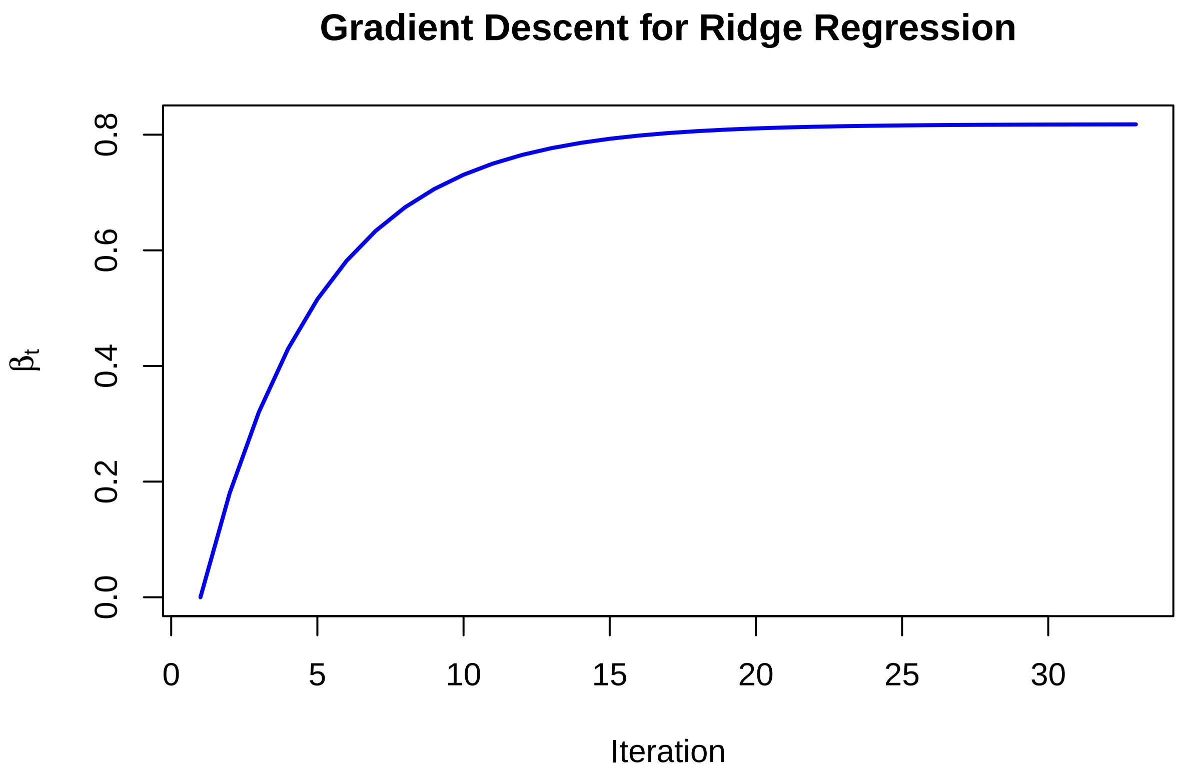
$\nabla f(\beta) = 2\sum_{i=1}^{n}(x_i\beta - y_i)x_i + 2\lambda\beta$

Initialize parameters for implementing gradient descent: The parameters to be initialized include:

$\lambda = 10$: Regularization parameter $\beta_0 = 0$: Initial value of $\beta$ $\alpha = 0.001$: Learning rate $\epsilon = 10^{-4}$: Convergence criterion

```r
# Data
X <- c(2, 4, 6, -2, -2, -6)
y <- c(2.5, 5.8, 3.7, -1.1, -3.7, -5)

# Parameters
lambda <- 10
beta <- 0
alpha <- 0.001
epsilon <- 10^-4

# To store beta values at each iteration
beta_history <- c(beta)

# Gradient descent loop
repeat {
  gradient <- 2 * sum(X * (X * beta - y)) + 2 * lambda * beta
  new_beta <- beta - alpha * gradient
  beta_history <- c(beta_history, new_beta)

  if (abs(new_beta - beta) < epsilon) {
    break
  }

  beta <- new_beta
}

# Ridge estimate
ridge_estimate <- beta
```

```r
# Plot the values of beta_t
plot(beta_history, type = "l", col = "blue", lwd = 2,
     xlab = "Iteration", ylab = expression(beta[t]),
     main = "Gradient Descent for Ridge Regression")
```




```r
cat("Ridge estimate:", ridge_estimate, "\n")
```

```
## Ridge estimate: 0.8178122
```

Question #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() function to print the results.

```r
library(ISLR2)
```

```
## Warning: package 'ISLR2' was built under R version 4.4.1
```

```r
Test <- ISLR2::Auto

model <- lm(mpg ~ . - name, data = Test)
summary(model)
```

```
##
## Call:
## lm(formula = mpg ~ . - name, data = Test)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -9.5903 -2.1565 -0.1169  1.8690 13.0604 
## 
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  -17.218435   4.644294  -3.707  0.00024 ***
## cylinders     -0.493376   0.323282  -1.526  0.12780    
## displacement   0.019896   0.007515   2.647  0.00844 ** 
## horsepower    -0.016951   0.013787  -1.230  0.21963    
## weight        -0.006474   0.000652  -9.929  < 2e-16 ***
## acceleration   0.080576   0.098845   0.815  0.41548    
## year           0.750773   0.050973  14.729  < 2e-16 ***
## origin         1.426141   0.278136   5.127 4.67e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 3.328 on 384 degrees of freedom
## Multiple R-squared:  0.8215, Adjusted R-squared:  0.8182 
## F-statistic: 252.4 on 7 and 384 DF,  p-value: < 2.2e-16
```

b) Is there a relationship between the predictors and the response?

Outputs:

F-Statistic: 252.4 (A high value suggests a strong relationship) p-value associated with the F-statistic: < 2.2e-16 (Indicates that it is very unlikely that all predictors have no relationship with mpg) The P-value is extremely low, so we reject the null hypothesis. Conclusion: Yes, there is a relationship between the predictors and the response variable mpg.

(c) For which of the predictors can you reject the null hypothesis $H_0 : \beta_j = 0$?

By looking at the P-values from the summary:

Displacement: p-value = 0.00844 Weight: p-value = 2e-16 Year: p-value = 2e-16 Origin: p-value = 4.67e-07 Conclusion: The predictors (Displacement, Weight, Year, and Origin) have p-values less than 0.05, meaning we reject the null hypothesis for these predictors. They are showing statistically significant relationships with mpg.

(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.

Explanation: A p-value of 0.03 indicates that if the hypothesis $H_0 : \beta_j = 0$ is true, there is a 3% chance of observing a test statistic as extreme as the one observed. It does not mean that the probability of $\beta_j = 0$ is 0.03.

Conclusion: No, it is incorrect to interpret a p-value of 0.03 as the probability that the null hypothesis is true. The p-value represents the probability of obtaining a test statistic at least as extreme as the one observed, assuming that the null hypothesis $H_0 : \beta_j = 0$ is true.