Ridge Regression Assignment Solution

by:Deepak Rai

1 Dataset

The given dataset is as follows:

X1	X2	X3	Y
100	25	10	20
200	35	12	30
300	45	15	40
400	55	17	50
500	65	19	60
600	75	22	70
700	85	25	80

2 Step 1: Perform Ordinary Least Squares (OLS) Regression

The general equation for Ordinary Least Squares (OLS) regression is:

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3$$

To calculate the coefficients β_0, β_1, β_2, and β_3, we apply standard linear regression. The design matrix X and response vector Y are:

$$X = \begin{pmatrix} 1 & 100 & 25 & 10 \\ 1 & 200 & 35 & 12 \\ 1 & 300 & 45 & 15 \\ 1 & 400 & 55 & 17 \\ 1 & 500 & 65 & 19 \\ 1 & 600 & 75 & 22 \\ 1 & 700 & 85 & 25 \end{pmatrix}$$

$$Y = \begin{pmatrix} 20 \\ 30 \\ 40 \\ 50 \\ 60 \\ 70 \\ 80 \end{pmatrix}$$

Using a statistical tool like Python or R, we can compute the OLS coefficients. The OLS equation is:

$$Y_{OLS} = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3$$

3 Step 2: Perform Ridge Regression

The Ridge Regression formula is:

$$\hat{\beta} = (X^T X + \lambda I)^{-1} X^T Y$$

We perform Ridge Regression using the same dataset with $\lambda = 5$.

4 Step 3: Compare Coefficients

The comparison of coefficients for OLS and Ridge Regression is shown below:

Model	β_0	β_1	β_2	β_3
OLS	(OLS Coefficient)	(OLS Coefficient)	(OLS Coefficient)	(OLS Coefficient)
Ridge ($\lambda = 5$)	(Ridge Coefficient)	(Ridge Coefficient)	(Ridge Coefficient)	(Ridge Coefficient)

As the regularization parameter λ increases, the Ridge coefficients shrink.

5 Step 4: Predictions for New Data

For a new set of values:

- $X_1 = 350$
- $X_2 = 50$
- $X_3 = 16$

The OLS prediction is:

$$Y_{OLS} = \beta_0 + \beta_1 \cdot 350 + \beta_2 \cdot 50 + \beta_3 \cdot 16$$

The Ridge prediction (for $\lambda = 5$) is:

$$Y_{Ridge} = \beta_0 + \beta_1 \cdot 350 + \beta_2 \cdot 50 + \beta_3 \cdot 16$$

The predicted values from both models are compared.

6 Step 5: Model Evaluation

We evaluate both models using the following metrics:

6.1 Mean Squared Error (MSE)

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

6.2 R-Squared

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

Where:

- SS_{res} is the sum of squared residuals.
- SS_{tot} is the total sum of squares.

7 Python Code Implementation

Here is the Python code to solve this problem using both OLS and Ridge Regression:

```
import numpy as np
import pandas as pd
from sklearn.linear_model import Ridge, LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

# Dataset
data = {
```

```
    ’X1’: [100, 200, 300, 400, 500, 600, 700],
    ’X2’: [25, 35, 45, 55, 65, 75, 85],
    ’X3’: [10, 12, 15, 17, 19, 22, 25],
    ’Y’: [20, 30, 40, 50, 60, 70, 80]
}

df = pd.DataFrame(data)

# Independent variables (X) and dependent variable (Y)
X = df[[’X1’, ’X2’, ’X3’]]
y = df[’Y’]

# Ordinary Least Squares (OLS) Regression
ols_model = LinearRegression()
ols_model.fit(X, y)
ols_predictions = ols_model.predict(X)

# Ridge Regression (=5)
ridge_model = Ridge(alpha=5)
ridge_model.fit(X, y)
ridge_predictions = ridge_model.predict(X)

# OLS coefficients
print("OLS Coefficients:", ols_model.coef_)

# Ridge coefficients
print("Ridge Coefficients (=5):", ridge_model.coef_)

# Mean Squared Error (MSE) and R-squared for OLS
ols_mse = mean_squared_error(y, ols_predictions)
ols_r2 = r2_score(y, ols_predictions)

# Mean Squared Error (MSE) and R-squared for Ridge
ridge_mse = mean_squared_error(y, ridge_predictions)
ridge_r2 = r2_score(y, ridge_predictions)

print(f"OLS MSE: {ols_mse}, OLS R-squared: {ols_r2}")
print(f"Ridge MSE: {ridge_mse}, Ridge R-squared: {ridge_r2}")

# Predictions for new data (X1=350, X2=50, X3=16)
new_data = np.array([[350, 50, 16]])
ols_new_pred = ols_model.predict(new_data)
ridge_new_pred = ridge_model.predict(new_data)

print(f"OLS Prediction for new data: {ols_new_pred}")
print(f"Ridge Prediction for new data: {ridge_new_pred}")
```

8 Conclusion

- Ridge Regression effectively shrinks the coefficients, making it useful when predictors are highly correlated or when overfitting is a concern.
- Depending on the value of λ, Ridge Regression may generalize better on unseen data, reducing overfitting.