report_2

Yi Yang

2024-11-22

Contents

#Assignment 2

0.1 Compute a linear regression fromm the training data.

Using `lm()`to fit a linear regression model without intercept, we got the summary of the model as follows:

0.2 Estimate traing and test MSE

The MSE for training and test data show as follows:

0.3 Commenting on Significant Variable Contributions to the Model.

The coefficient`Jitter.Abs.`,`Shimmer.APQ5`,`Shimmer.APQ11`,`NHR`,`HNR`,`DFA`,`PPE` is highly significant, as their P-value are less than 0.001,which suggests that they have a strong and statistically reliable impact on the response variable. The coefficient `Shimmer` is statistically significant,with a P-value < 0.01,indicating that `Shimmer` has a meaningful and reliable impact on the response variable.

0.4 Using `RidgeOpt`to compute optimal parameters when lambda=1, 100 and 1000.

The training and test MSE for different lambda and the degrees of freedom shows as follws:

0.5 Comenting on which is the most appropriate penalty parameter.

Among the selected penalty parameters, lambda = 100 is the most appropriate .It provides the lowest test data MSE(0.9323316). Besides, the degree of freedom for lambda = 100 is 9.924887, which shows a balanced model complexity that avoids both overfitting and underfitting.

1 Assignment 4. Theory

1.1 Express the cost fucntion of the linear regression in the matrix form.

The cost function for the linear regression in the matrix form is defined as:

$$J(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}(x_i;\theta) - y_i\right)^2 = \frac{1}{n}\|\hat{y} - y\|_2^2 = \frac{1}{n}\|X\theta - y\|_2^2 = \frac{1}{n}\|\epsilon\|_2^2$$

where: - $\hat{y} = X\theta$ is the vector of the predicted values - y is the vector of true values - $\epsilon = y - \hat{y}$ is the residuals - $\|\cdot\|$ denotes the Euclidean norm This cost function is also referred to as the least squares cost. It can be found on page 40,formula(3.11).

1.2 Code for assignment 2