GATE

Machine Learning

Regression

DPP: 2

Q1 Consider the linear regression model $Y = X\beta + \varepsilon$ with $\varepsilon \sim N$ (0n, σε^2 Inn). This model (without intercept) is fitted to data using the ridge regression estimator β^(λ) = arg minβ ||Y − Xβ|| 2 2 + λ||β|| 2 2 with λ > 0.

The data are:

X^T =(−1 1 1 −1) and Y^T = (−1.5 2.9 −3.5 0.7)

What is the maximum likelihood/ordinary least squares estimator of the regression parameter for λ =0?

(A) [-0.3, 0.05]
(B) [-0.5, 0.1]
(C) [0.1, -0.2]
(D) [0.05, -0.3]

Q2 Suppose you are training a Ridge Regression model for a particular task and notice the following training error and validation RSS

Train: 57

Validation: 32,714

Would your next to try a Ridge model with a larger or smaller λ

(A) Larger
(B) Smaller
(C) λ does not have an effect here
(D) Neither larger nor smaller

Q3 How does ridge regression help in dealing with overfitting in a dataset with a large number of predictors?

(A) By increasing the number of observations
(B) By introducing a penalty term that shrinks the coefficients towards zero
(C) By removing outliers from the dataset
(D) By reducing the number of predictors

Q4 What role does the Regularization parameter (λ) play in controlling the bias-variance trade-off in a ridge regression model?

(A) λ controls the number of predictors in the model
(B) λ controls the degree of multicollinearity among predictors
(C) λ balances the trade-off between bias and variance
(D) λ has no effect on the model's performance

Q5 What are the ridge regression coefficients for a dataset with predictors X1, X2, X3, and response variable Y, using a regularization parameter (λ) value of 0.5?

(A) It depends on the number of observations
(B) They are calculated using the formula: β^ridge = (X^T X + λ I)^-1 X^T Y
(C) Ridge regression does not provide coefficients
(D) They are the same as OLS regression coefficients

Q6 Increasing the regularizing coefficient value for a ridge regressor will

(i) Increase or maintain model bias.
(ii) Decrease model bias.
(iii) Increase or maintain model variance.
Decrease model variance

(A) i & ii
(B) i & iv
(C) ii & iii
(D) ii & iv

Q7 Using the data X=[-3,5,4] and Y =[-10,20,20], assuming a ridge penalty λ = 50, what ratio versus the Maximum Likelihood Estimate (MLE) estimate w_{mle} do you think the ridge regression L2 estimate estimate w_{ridge} estimate will be?

(A) 2
(B) 1
(C) 0.6
(D) 0.5

Q8 As the regularization parameter increases in Ridge regression, do the regression coefficients decrease?

Q9 Which of the following statements are true?

Statement 1: Modifying the cost function can be done by incorporating a penalty equal to the square of the coefficients' magnitudes.

Statement 2: Ridge and Lasso regression are among the basic methods used to mitigate model complexity and counter overfitting issues, which can arise in simple linear regression

(A) Statement 1 is incorrect, and statement 2 is correct.

(B) Statement 1 is correct, and statement 2 is incorrect.

(C) Both statements 1 and 2 are correct.

(D) Both statements 1 and 2 are incorrect.

Answer Key

Q1	(A)
Q2	(A)
Q3	(B)
Q4	(C)
Q5	(B)
Q6	(B)
Q7	(D)
Q8	true
Q9	(C)

Hints & Solutions

Q1 Text Solution:

$\hat{\beta}_{OLS} = (X^T X)^{-1} X^T Y$

After Calculation:

$$X^T X = \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix}$$

$$(X^T X)^{-1} = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$$

$$X^T Y = \begin{pmatrix} -1.5 + 2.9 + 3.5 + 0.7 \\ -1.5 - 2.9 - 3.5 - 0.7 \end{pmatrix} = \begin{pmatrix} 5.6 \\ -8.6 \end{pmatrix}$$

So,

$$\hat{\beta}_{OLS} = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix} * \begin{pmatrix} 5.6 \\ -8.6 \end{pmatrix} = \begin{pmatrix} -0.3 \\ 0.05 \end{pmatrix}$$

Thus, the greatest likelihood/ordinary slightest squares estimator of the relapse parameter for λ = is [-0.3, 0.05].

Q2 Text Solution:

Since the approval mistake is essentially higher than the preparing blunder, it recommends that the show may be overfitting. To address overfitting in Edge Relapse, we have to increment the regularization quality by expanding λ. This will offer assistance to anticipate overfitting by penalising expansive coefficient values more intensely, driving to a less difficult and more generalised show.

Q3 Text Solution:

Edge relapse makes a difference in managing with overfitting in datasets with a huge number of indicators by presenting a punishment term (λ) that penalizes the size of coefficients. This punishment term shrivels the coefficients towards zero, viably diminishing their affect on the demonstrate. By contracting the coefficients, edge relapse avoids the demonstrate from fitting commotion within the information as well closely, hence moving forward its capacity to generalize to concealed information and lessening overfitting.

Q4 Text Solution:

In edge relapse, the regularization parameter (λ) controls the punishment connected to the coefficients. By expanding λ, the demonstrate penalizes huge coefficient values, which makes a difference in diminishing the fluctuation of the show. Be that as it may, it too increments the predisposition of the demonstrate by somewhat underfitting the information. Hence, λ acts as a adjusting parameter that controls the trade-off between inclination and change. Altering λ permits us to discover an ideal adjust that minimizes both predisposition and change, in this way progressing the in general execution of the show.

Q5 Text Solution:

Edge relapse coefficients are not the same as OLS relapse coefficients. They're calculated employing a particular equation that consolidates the regularization parameter (λ), the indicator lattice (X), and the reaction variable (Y). This equation alters the coefficients to account for regularization, guaranteeing way better generalization and control over show complexity.

Q6 Text Solution:

Regularization adds a penalty term to the loss function, which helps in reducing the complexity of the model by shrinking the coefficients towards zero. This tends to reduce the flexibility of the model, leading to higher bias. So, increasing the regularization coefficient will either increase or maintain model bias. Regularization reduces the variance of the model by shrinking the coefficients towards zero, which helps in preventing overfitting. As the regularization coefficient increases, the penalty for large coefficients increases, leading to a more constrained model with lower variance.

Q7 Text Solution:

$$X^TX = \begin{bmatrix} (-3)^2 & -3*5 & -3*4 \\ -3*5 & 5^2 & 5*4 \\ -3*4 & 5*4 & 4^2 \end{bmatrix}$$

$$= \begin{bmatrix} 50 & -15 & -12 \\ -15 & 50 & 20 \\ -12 & 20 & 41 \end{bmatrix}$$

$$X^TX = \begin{bmatrix} -3 \\ 5 \\ 4 \end{bmatrix} \begin{bmatrix} -10 & 20 & 20 \end{bmatrix} = \begin{bmatrix} 30 \\ 80 \\ 80 \end{bmatrix}$$

$w_{ridge} = (X^TX+\lambda I)^{-1}X^TY$

$$= \left(\begin{bmatrix} 50 & -15 & -12 \\ -15 & 50 & 20 \\ -12 & 20 & 41 \end{bmatrix} + 50 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 30 \\ 80 \\ 80 \end{bmatrix}$$

= [0.9,0.8,0.6]

$w_{mle} = (X^TX)^{-1}X^TY$

$$= \begin{bmatrix} 50 & -15 & -12 \\ -15 & 50 & 20 \\ -12 & 20 & 41 \end{bmatrix}^{-1} \begin{bmatrix} 30 \\ 80 \\ 80 \end{bmatrix}$$

= [1.8,1.6,1.2]

W_{ridge} / w_{mle} = [0.5,0.5,0.5] = 0.5

Q8 Text Solution:

In Ridge regression, as the regularization parameter λ increases, the penalty term added to the loss function becomes stronger. This penalty term is proportional to the sum of the squares of the coefficients. Consequently, to minimize the total loss function (which includes both the original loss and the penalty term), the coefficient estimates are shrunk towards zero.

Mathematically, the Ridge regression loss function can be represented as:

Loss=OLS Loss+$\lambda \sum^{p}_{j=1} \beta_j^2$

As λ increases, the penalty term becomes more dominant, and the minimization of the loss function requires smaller coefficient estimates.

Q9 Text Solution:

Statement 1 correctly identifies that modifying the cost function in regression models can be achieved by adding a penalty term equivalent to the square of the coefficients' magnitudes. This penalty term is commonly used in Ridge regression to control the size of the coefficients and prevent overfitting.

Statement 2 accurately states that Ridge and Lasso regression are indeed basic techniques employed to address model complexity and mitigate overfitting issues that may arise in simple linear regression. Both Ridge and Lasso regression methods introduce regularization to the model, which helps to constrain the coefficients and prevent them from becoming too large, thus reducing overfitting.