Graded

This document has 10 questions.

Question-1

Statement

Assume that for a certain linear regression problem involving 4 features, the following weight vectors produce an equal amount of mean square error:

w_1 = [2, 2, 3, 1]

w_2 = [1, 1, 3, 1]

w_3 = [3, 2, 4, 1]

w_4 = [1, 2, 1, 1]

Which of the weight vector is likely to be chosen by ridge regression?

Options

(a)

w_1

(b)

w_2

(c)

w_3

(d)

w_4

Answer

(d)

Solution

Total error = MSE + $\lambda||w||^2$

If MSE for all the given weights is same, the weight vector whose squared length is the least will be chosen by Ridge Regression.

Question-2

Statement

Assuming that in the constrained version of ridge regression optimization problem, following are the weight vectors to be considered, along with the mean squared error (MSE) produced by each:

w_1 = [2, 2, 3, 1], MSE = 3

w_2 = [1, 1, 3, 1], MSE = 5

w_3 = [3, 2, 4, 1], MSE = 8

w_4 = [1, 2, 1, 1], MSE = 9

If the value of θ is 13, which of the following weight vectors will be selected as the final weight vector by ridge regression?

Note: θ is as per lectures. That is, $||w||^2 \leq \theta$

Options

(a)

w_1

(b)

w_2

(c)

w_3

(d)

w_4

Answer

(b)

Solution

We need to minimize MSE such that $||w||^2 \leq \theta$

$||w_1||^2 = 18, ||w_2||^2 = 12, ||w_3||^2 = 30, ||w_4||^2 = 7$

$||w||^2 \leq 13$ for w_2 and w_4.

However, the MSE for w_2 is lesser than w_4.

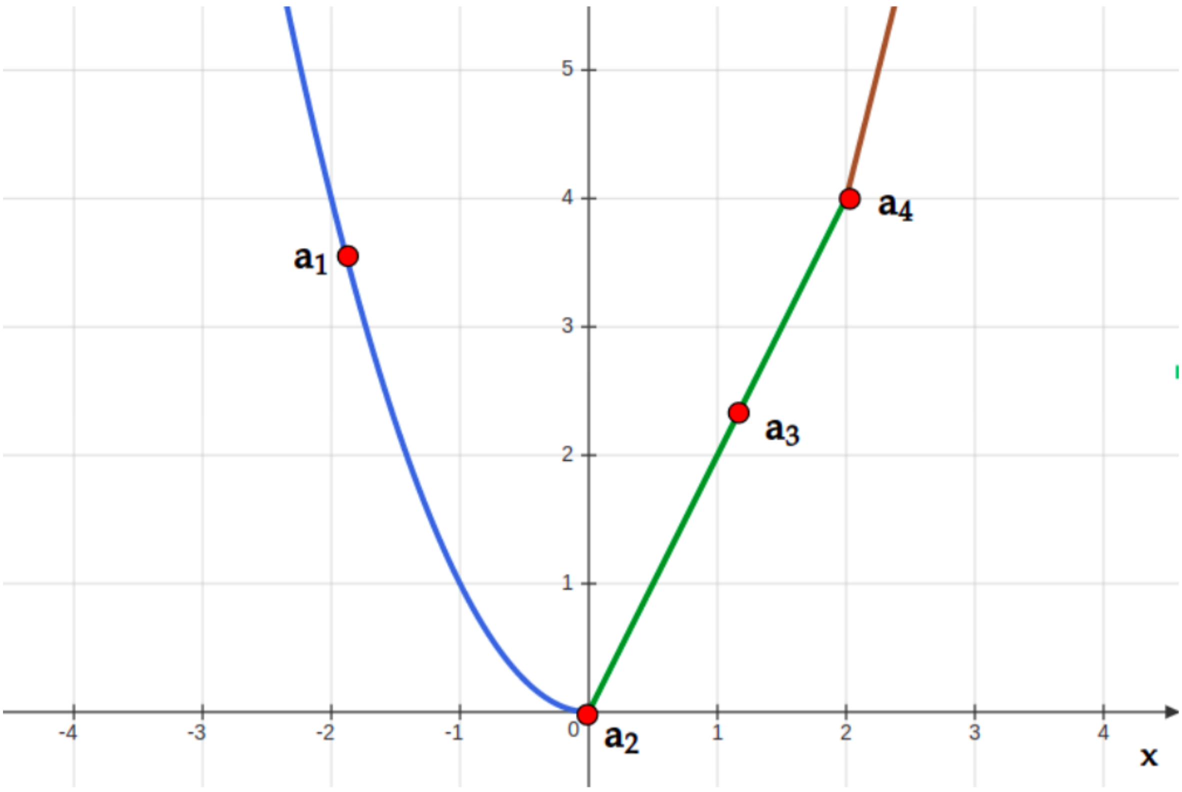
Hence, w_2 will be chosen.

Question-3

Statement

Consider the following piece-wise function as shown in the image:

$$y(x) = \begin{cases} x^2 & x \leq 0 \\ 2x & 0 \leq x \leq 2 \\ 4x - 4 & 2 \leq x \end{cases}$$



How many sub-gradients are possible at points a_1, a_2, a_3 and a_4?

Options

(a)

$a_1 : Many, a_2 : One, a_3 : Many, a_4 : One$

(b)

$a_1 : One, a_2 : Many, a_3 : Many, a_4 : One$

(c)

$a_1 : One, a_2 : Many, a_3 : One, a_4 : Many$

(d)

$a_1 : Many, a_2 : One, a_3 : One, a_4 : Many$

Answer

(c)

Solution

$a_1$lies on the part of the function which is differentiable. For a differentiable function (subpart), only one sub-gradient is possible which is the gradient itself.

a_2 lies at the intersection of two x^2 and $2x$. The function is not differentiable at this point (as left slope is different from right slope). Hence there are multiple sub-gradients possible at a_2.

a_3 lies on the part of the function which is differentiable. For a differentiable function (subpart), only one sub-gradient is possible which is the gradient itself.

a_4 lies at the intersection of two $2x$ and $4x - 4$. The function is not differentiable at this point (as left slope is different from right slope). Hence there are multiple sub-gradients possible at a_2

Question-4

Statement

For a data set with 1000 data points and 50 features, 10-fold cross-validation will perform validation of how many models?

Options

(a)

10

(b)

50

(c)

1000

(d)

500

Answer

(a)

Solution

In 10-fold cross validation, the data will be divided into 10 parts. In each of ten iterations, a model will be built using nine of these parts and the remaining part will be used for validation. Hence, in total, ten models will be validated.

Question-5

Statement

For a data set with 1000 data points and 50 features, assume that you keep 80% of the data for training and remaining 20% of the data for validation during k-fold cross-validation. How many models will be validated during cross-validation?

Options

(a)

80

(b)

20

(c)

5

(d)

4

Answer

(c)

Solution

If 20% of the data is used for validation, that means, 1/5th part is used for validation, which means, 5-fold cross validation is being performed. In each iteration, one model will be validated. Hence, total 5 models will be validated.

Question-6

Statement

For a data set with 1000 data points and 50 features, how many models will be trained during Leave-One-Out cross-validation?

Options

(a)

1000

(b)

50

(c)

5000

(d)

20

Answer

(a)

Solution

In leave one out cross-validation, only one data point is used for validation in each iteration, and the remaining n-1 data points are used for training. Hence a total of n = 1000 models will be trained.

Question-7

Statement

The mean squared error of $\hat{w}_{ML}$ will be small if

Options

(a)

The eigenvalues of XX^T are small.

(b)

The eigenvalues of $(XX^T)^{-1}$ are large.

(c)

The eigenvalues of XX^T are large.

(d)

The eigenvalues of $(XX^T)^{-1}$ are small.

Answer

(c), (d)

Solution

Mean Squared error of $\hat{w}_{ML} = \sigma^2 trace(XX^T)^{-1}$. Trace of a matrix = sum of eigenvalues.

If the eigenvalues of XX^T are large, the eigenvalues of $(XX^T)^{-1}$ will be small. Hence, trace will be small and in turn MSE will be small.

Question-8

Statement

The eigenvalues of a 3×3 matrix A are 2, 5 and 1. What will be the eigenvalues of the matrix A^{-1}

Options

(a)

4, 25, 1

(b)

2, 5, 1

(c)

0.5, 0.2, 1

(d)

0.6, 0.9, 0.1

Answer

(c)

Solution

If the eigenvalues of A are a, b and c, then the eigenvalues of A^{-1} will be 1/a, 1/b and 1/c.