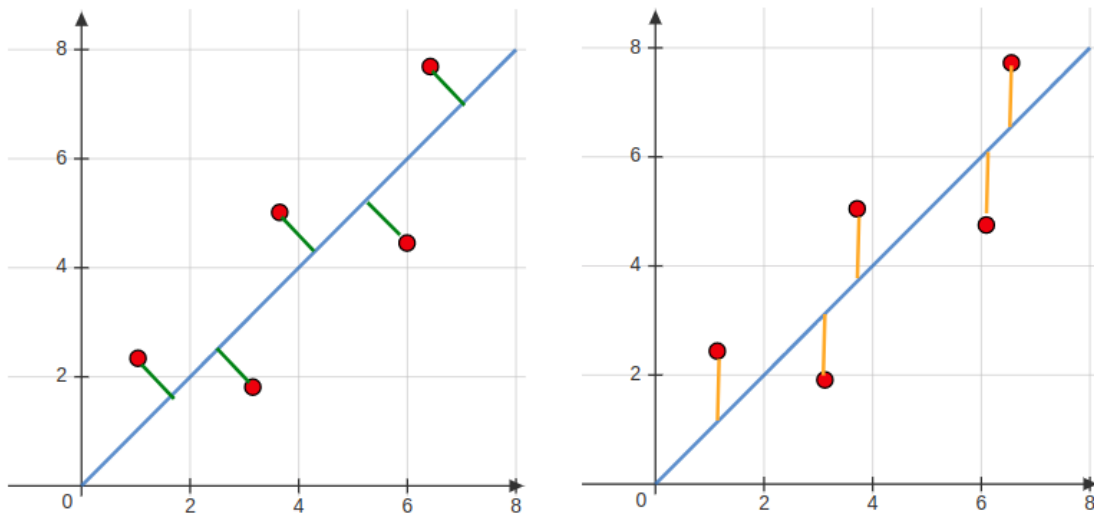# Practice

## Question-1

### Statement

Consider the following image:



The images shows two types of lines: Green and Orange. Which of these lines are used for calculating the residuals in linear regression?

### Options

**(a)**

Green

**(b)**

Orange

**(c)**

Does not matter; Any of these may be used.

**(d)**

None of these

### Answer

(b)

## Solution

In linear regression, the task is to predict the y-axis value given the x-axis value (for a one variable scenario). The regression line tends to fit the given data points as best as possible, such that the residual error may be minimized. Here the residual error is the distance between the actual y-value and the y-value that is predicted by the regression line. Hence, this distance line will be a vertical line. Hence the orange lines will represent these residuals.

The green lines on the other hand, represent the errors in PCA, as in PCA, the principal components tend to fit the data points such that the variance in the data is captured. The errors in PCA, thus are perpendicular lines.

# Question-2

## Statement

Assume that the eigenvalues of a $3 \times 3$ matrix $A$ are 2, 3, 1. What will be the eigenvalues of the matrix $A + 5I$ where $I$ is a $3 \times 3$ identity matrix.

## Options

**(a)**

2, 3, 1

**(b)**

7, 8, 6

**(c)**

10, 15, 5

**(d)**

6, 6, 6

## Answer

(b)

## Solution

If a, b, c are eigenvalues of a matrix A, then then the eigenvalues of the matrix $A + \alpha I$ will be $a + \alpha, b + \alpha$ and $c + \alpha$.

# Question-3

## Statement

What will be the trace of $(XX^T + \lambda I)^{-1}$?

Note: $\lambda_i$ are the eigenvalues of $XX^T$

## Options

**(a)**

$\sum_{i=1}^{d} \lambda_i + \lambda$

**(b)**

$\sum_{i=1}^{d} 1 + \lambda_i$

**(c)**

$\sum_{i=1}^{d} \dfrac{1}{1 + \lambda_i}$

**(d)**

$\sum_{i=1}^{d} \dfrac{1}{\lambda_i + \lambda}$

## Answer

(d)

## Solution

The trace of a matrix is the sum of its eigenvalues.

If $\lambda_i$'s are the eigenvalues of $XX^T$, then $\lambda_i + \lambda$ will be the eigenvalues of $XX^T + \lambda I$.

If a, b are eigenvalues of a matrix M, then 1/a and 1/b will be the eigenvalues of the matrix $M^{-1}$.

So, the eigenvalues of $(XX^T + \lambda I)^{-1}$ will be $\dfrac{1}{\lambda_i + \lambda}$.

Hence the trace will be $\sum_{i=1}^{d} \dfrac{1}{\lambda_i + \lambda}$

# Question-4

## Statement

Assume that as part of 3-fold cross-validation, the data was divided into parts $f_1, f_2, f_3$. Subsequently, cross-validation was performed. In this context, which of the following is True?

## Options

### (a)

In each iteration of cross-validation, one of $f_1, f_2, f_3$ will be used for training, and the remaining two parts will be used for testing.

### (b)

In each iteration of cross-validation, two of $f_1, f_2, f_3$ will be used for training, and the remaining one part will be used for validation.

### (c)

In each iteration of cross-validation, two of $f_1, f_2, f_3$ will be used for training, and the remaining one part will be used for testing.

### (d)

In each iteration of cross-validation, one of $f_1, f_2, f_3$ will be used for training, and the remaining two parts will be used for validation.

## Answer

(b)

## Solution

In k-fold cross-validation, the training data is split into k parts.

K-fold cross validation involves k iterations.

In each iteration, one of the k parts is used for validation and the remaining k-1 parts are used for training.

Hence, in k-fold cross validation, one part will be used for validation and remaining two will be used for training.

# Question-5

## Statement

Assume that we obtain the weight vector $w = [4, 6, 8, 10]$ when $\lambda$= 3 is used in Ridge Regression.

If $\lambda$ is increased to 6, which of the following is most likely to be the updated weight vector?

## Options

**(a)**

$w = [7, 9, 11, 13]$

**(b)**

$w = [1, 3, 6, 9]$

**(c)**

$w = [10, 12, 13, 15]$

## Answer

(b)

## Solution

In regularization, as the value of $\lambda$ (which represents penalty for higher weights) is increased, the weights are further reduced.

Therefore, in the given question, as the value of $\lambda$ is increased from 3 to 6, the penalty for having higher weights will increase, thus shrinking the weights.

# Question-6

## Statement

The mean squared error of $\hat{w}_{ML}$ could come out to be large if

## Options

**(a)**

Trace of $XX^T$ is large.

**(b)**

Trace of $(XX^T)^{-1}$ is large.

**(c)**

$\sigma^2$ is large.

**(d)**

$\lambda_i$'s are large.

## Answer

(b), (c)

## Solution

MSE = $\sigma^2 trace(XX^T)^{-1}$

Hence, MSE will come out to be large if $\sigma^2$ is large or trace of $(XX^T)^{-1}$ is large.

# Question-7

## Statement

Cross validation may be useful as it may help in

## Options

### (a)

Reducing the trace of the inverse of covariance matrix.

### (B)

Increasing the trace of the inverse of covariance matrix.

## Answer

(a)

## Solution

The expression for $w^*$ is given by $(XX^T)^{-1}Xy$

Cross validation results into a weight vector $(XX^T + \lambda I)^{-1}Xy$

This in turn reduces the trace of the inverse of covariance matrix, thus reducing the MSE.

# Question-8

## Statement

Assume that Lasso is applied to a data set containing only one feature. If $\lambda$ is increased to a high value, which of the following is correct?

## Options

**(a)**

The regression line may become horizontal.

**(b)**

The regression line may become horizontal.

**(c)**

The regression line will remain the same.

### Answer

(a)

## Solution

In lasso, as the value of $\lambda$ is increased, due to a very high penalty, the coefficient of the only feature becomes zero.

The coefficient represents the slope. And when slope = 0, it results into a horizontal line.

# Question-9

## Statement

In Ridge Regression, if $\lambda$ tends to zero, the solution approaches to

## Options

**(a)**

Zero

**(b)**

One

**(c)**

Linear Regression

**(d)**

Lasso Regression

**(e)**

Infinity

## Answer

(c)

## Solution

The weight vector expression for linear regression is $(XX^T)^{-1}Xy$

The weight vector expression for ridge regression is $(XX^T + \lambda I)^{-1}Xy$

When $\lambda$ tends to zero, the solution of ridge is same as that of linear regression.