Linear Regression Interview Questions & Answers

Q1. True-False: Linear Regression is a supervised machine learning algorithm.

A) TRUE

B) FALSE

Q2. True-False: Linear Regression is mainly used for Regression.

A) TRUE

B) FALSE

Q3. True-False: It is possible to design a Linear regression algorithm using a neural network.

A) TRUE

B) FALSE

Q4. Which of the following methods do we use to find the best-fit line for data in Linear Regression?

A) Least Square Error

B) Maximum Likelihood

C) Logarithmic Loss

D) Both A and B

Q5. Which of the following evaluation metrics can be used to evaluate a model while modeling a continuous output variable?

A) AUC-ROC

B) Accuracy

C) Logloss

D) Mean-Squared-Error

Q6. True-False: Lasso Regularization can be used for variable selection in Linear Regression.

A) TRUE

B) FALSE

Q7. Which of the following is true about residuals?

A) Lower is better

B) Higher is better

C) A or B depending on the situation

D) None of these

Q8. Suppose we have N independent variables (X1, X2… Xn) and Y's dependent variable.

Now Imagine that you are applying linear regression by fitting the best-fit line using the least square error on this data. You found that the correlation coefficient for one of its variables (Say X1) with Y is -0.95.

Which of the following is true for X1?

A) Relation between the X1 and Y is weak

B) Relation between the X1 and Y is strong

C) Relation between the X1 and Y is neutral

D) Correlation can't judge the relationship

Q9. Looking at the above two characteristics, which of the following options is the correct Pearson correlation between V1 and V2?

If you are given the two variables V1 and V2, which follow the below two characteristics:

1. If V1 increases, then V2 also increases

2. If V1 decreases, then V2 behavior is unknown

A) Pearson correlation will be close to 1

B) Pearson correlation will be close to -1

C) Pearson correlation will be close to 0

D) None of these

Q10. True- False: Overfitting is more likely when you have a huge amount of data to train.

A) TRUE

B) FALSE

Q11. We can compute the coefficient of linear regression with the help of an analytical method called "Normal Equation." Which of the following is/are true about Normal Equations?

    1- We don't have to choose the learning rate.

    2- It becomes slow when the number of features is very large.

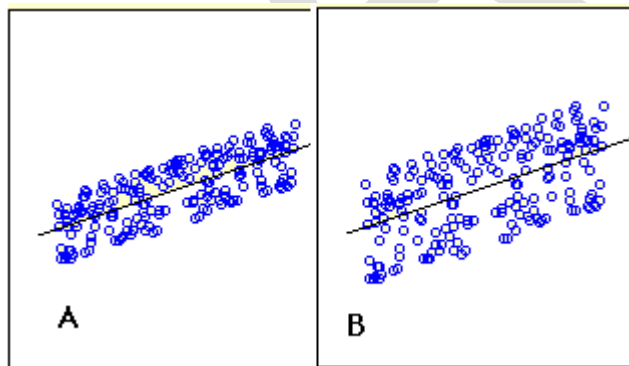   3- There is no need to iterate.

A) 1 and 2

B) 1 and 3

C) 2 and 3

D) 1,2 and 3

Q12. Which of the following statement is true about the sum of residuals of A and B?

Below graphs show two fitted regression lines (A & B) on randomly generated data. Now, I want to find the sum of residuals in both cases, A and B.

Note:

Scale is the same in both graphs for both axes.

X-axis is the independent variable, and Y-axis is the dependent variable.

A) A has a higher sum of residuals than B

B) A has a lower sum of residual than B

C) Both have the same sum of residuals

D) None of these

Suppose you have fitted a complex regression model on a dataset. Now, you are using Ridge regression with penalty x.

Q13. Choose the option which describes bias in the best manner.

A) In the case of a very large x, bias is low

B) In the case of a very large x, bias is high

C) We can't say about bias

D) None of these

Q14. What will happen when you apply a very large penalty?

A) Some of the coefficients will become absolute zero

B) Some of the coefficients will approach zero but not absolute zero

C) Both A and B depending on the situation

D) None of these

Q15. What will happen when you apply a very large penalty in the case of Lasso regression?

A) Some of the coefficients will become zero

B) Some of the coefficients will be approaching zero but not absolute zero

C) Both A and B depending on the situation

D) None of these

Q16. Which of the following statement is true about outliers in Linear regression?

A) Linear regression is sensitive to outliers

B) Linear regression is not sensitive to outliers

C) Can't say

D) None of these

Q17. Suppose you plotted a scatter plot between the residuals and predicted values in linear regression and found a relationship between them. Which of the following conclusion do you make about this situation?

A) Since there is a relationship means our model is not good

B) Since there is a relationship means our model is good

C) Can't say

D) None of these

Suppose that you have a dataset D1 and you design a linear model of degree 3 polynomial and find that the training and testing error is "0" or, in other words, it perfectly fits the data.

Q18. What will happen when you fit a degree 4 polynomial in linear regression?

A) There is a high chance that degree 4 polynomial will overfit the data

B) There is a high chance that degree 4 polynomial will underfit the data

C) Can't say

D) None of these

Q19. What will happen when you fit a degree 2 polynomial in linear regression?

A) It is a high chance that degree 2 polynomial will overfit the data

B) It is a high chance that degree 2 polynomial will underfit the data

C) Can't say

D) None of these

Q20. In terms of bias and variance. Which of the following is true when you fit degree 2 polynomial?

A) Bias will be high, and variance will be high

B) Bias will be low, and variance will be high

C) Bias will be high, and variance will be low

D) Bias will be low, and variance will be low

Below are three graphs, A, B, and C, between the cost function and the number of iterations, I1, I2, and I3, respectively.
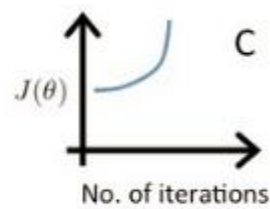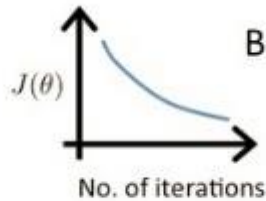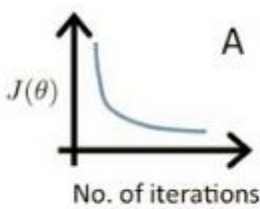
Q21. Suppose l1, l2, and l3 are the three learning rates for A, B, and C, respectively. Which of the following is true about l1,l2, and l3?

A) l2 < l1 < l3

B) l1 > l2 > l3

C) l1 = l2 = l3

D) None of these

We have been given a dataset with n records in which we have an input attribute as x and an output attribute as y. Suppose we use a linear regression method to model this data. To test our linear regressor, we split the data in the training set and test a set randomly.

Q22. Now we increase the training set size gradually. As the training set size increases, what do you expect will happen with the mean training error?

A) Increase

B) Decrease

C) Remain constant

D) Can't Say

Q23. What do you expect will happen with bias and variance as you increase the size of training data?

A) Bias increases, and Variance increases

B) Bias decreases, and Variance increases

C) Bias decreases, and Variance decreases

D) Bias increases, and Variance decreases

E) Can't Say False

Consider the following data where one input(X) and one output(Y) are given.
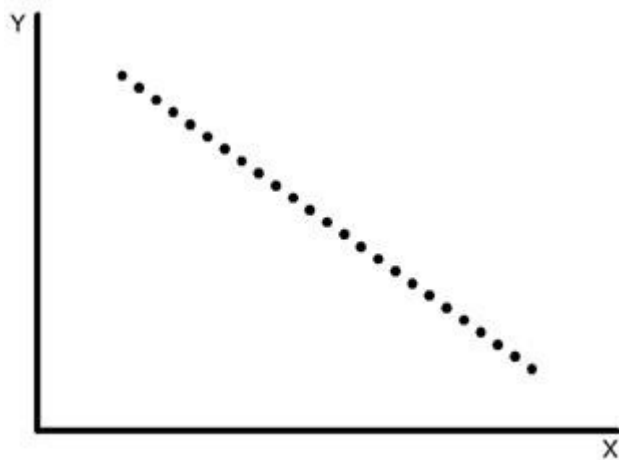
Q24. What would be the root mean square training error for this data if you run a Linear Regression model of the form (Y = A0+A1X)?

A) Less than 0

B) Greater than zero

C) Equal to 0

D) None of these

Suppose you have been given the following scenario for training and validation error for Linear Regression.

Scenario	Learning Rate	Number of iterations	Training Error	Validation Error
1	0.1	1000	100	110
2	0.2	600	90	105
3	0.3	400	110	110
4	0.4	300	120	130
5	0.4	250	130	150

Q25. Which of the following scenario would give you the right hyperparameter?

A) 1

B) 2

C) 3

D) 4

Q26. Suppose you got the tuned hyperparameters from the previous question. Now, Imagine you want to add a variable in variable space such that this added feature is important.

Which of the following thing would you observe in such a case?

A) Training Error will decrease, and Validation error will increase

B) Training Error will increase, and Validation error will increase

C) Training Error will increase, and Validation error will decrease

D) Training Error will decrease, and Validation error will decrease

E) None of the above

Suppose you got a situation where you find that your linear regression model is underfitting the data.

Q27. In such a situation, which of the following options would you consider?

Add more variables

Start introducing polynomial degree variables

Remove some variables

A) 1 and 2

B) 2 and 3

C) 1 and 3

D) 1, 2 and 3

Q28. Now the situation is the same as written in the previous question (under-fitting). Which of the following regularization algorithms would you prefer?

A) L1

B) L2

C) Any

D) None of these

We have this data set for number of hours and the exam test regarding for number of hours

Hours	Exam Result
1	2
2	4
3	6
4	8
5	10

Assume the model is $y = wx + b$.

What are the optimal values for w (slope) and b (intercept) after training the model?

A) $w = 2, b = 0$

B) $w = 1, b = 1$

C) $w = 3, b = -1$

D) $w = 0, b = 2$

If the current parameters are $w = 0$ and $b = 0$, calculate the gradient for w and b for the first data point $(X = 1, y = 2)$.

$$dw = \frac{-2}{n} \sum_{i=1}^{n} X_i(y_i - y_{\text{pred}})$$

$$db = \frac{-2}{n} \sum_{i=1}^{n} (y_i - y_{\text{pred}})$$

A) $dw = -4, db = -2$

B) $dw = -2, db = -1$

C) $dw = -6, db = -3$

D) $dw = -8, db = -4$

In Gradient Descent, the weights w and b are updated using the formulas:

$$w = w - \alpha \cdot \mathrm{dw}, \quad b = b - \alpha \cdot \mathrm{db}$$

If $w = 0, b = 0, \alpha = 0.01$, and the gradients are $\mathrm{dw} = -4$ and $\mathrm{db} = -2$, what are the updated values of w and b?

A) $w = 0.04, b = 0.02$
B) $w = -0.04, b = -0.02$
C) $w = 0.01, b = 0.005$
D) $w = 0.08, b = 0.04$

The Mean Squared Error (MSE) is given by:

$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - y_{\mathrm{pred}})^2$$

For $X = [1, 2, 3, 4, 5]$, $y = [2, 4, 6, 8, 10]$, and predictions $y_{\mathrm{pred}} = [1.8, 3.6, 5.4, 7.2, 9.0]$, calculate the MSE.

A) 0.1
B) 0.2
C) 0.4
D) 0.8

Why is a small learning rate (α) preferred over a large one in Gradient Descent?

A) To reach the global minimum more quickly
B) To avoid overshooting the minimum
C) To prevent underfitting
D) To guarantee faster updates

L1----L2

Lasso Regression is a type of linear regression that uses L1 regularization to penalize the magnitude of coefficients in the model. The Lasso objective function is:

$$Lasso\ Loss = \frac{1}{2m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{n} |w_j|$$

Suppose we have below small dataset for example

X1	X2	Y
1	2	5
2	3	7
3	4	9
4	5	11

Step 1: Build the Linear Model

$$y = w_0 + w_1 x_1 + w_2 x_2$$

Step 2: Loss Function for Lasso

For Lasso regression, the objective function is:

$$Lasso\ Loss = \frac{1}{2m} \sum_{i=1}^{m} (y_i - (w_0 + w_1 x_{i1} + w_2 x_{i2}))^2 + \lambda(|w_1| + |w_2|)$$

Suppose $\lambda = 0.1$

Step 3: Initial Predictions

Let's assume initial values for the weights:

$$w_0 = 1, \quad w_1 = 0.5, \quad w_2 = 1$$

Now, we can calculate the predicted values for y (denoted as $\hat{y}$):

$$\hat{y}_1 = w_0 + w_1 x_{11} + w_2 x_{12} = 1 + 0.5 \cdot 1 + 1 \cdot 2 = 3.5$$

$$\hat{y}_2 = w_0 + w_1 x_{21} + w_2 x_{22} = 1 + 0.5 \cdot 2 + 1 \cdot 3 = 4.5$$

$$\hat{y}_3 = w_0 + w_1 x_{31} + w_2 x_{32} = 1 + 0.5 \cdot 3 + 1 \cdot 4 = 6.5$$

$$\hat{y}_4 = w_0 + w_1 x_{41} + w_2 x_{42} = 1 + 0.5 \cdot 4 + 1 \cdot 5 = 7.5$$

Predictions: $\hat{y} = [3.5, 4.5, 6.5, 7.5]$

Step 4: Compute the Squared Error

Now calculate the squared errors for each prediction:

$$(5 - 3.5)^2 = 2.25$$

$$(7 - 4.5)^2 = 6.25$$

$$(9 - 6.5)^2 = 6.25$$

$$(11 - 7.5)^2 = 12.25$$

Total squared error:

$$\frac{1}{4}(2.25 + 6.25 + 6.25 + 12.25) = 6$$

Step 5: Add the Regularization Term

Now compute the L1 regularization term. Since $\lambda = 0.1$:

$$L1\,Term = 0.1 \cdot (|w_1| + |w_2|) = 0.1 \cdot (|0.5| + |1|) = 0.1 \cdot (0.5 + 1) = 0.15$$

Step 6: Total Loss Function

Now we can compute the total loss function:

$$Total\ Loss = 6 + 0.15 = 6.15$$

Step 7: Update the Weights (Gradient Descent)

1. **What is the primary goal of Lasso regression?**

   - A) Minimize the sum of squared errors only.

   - B) Minimize the sum of squared errors with an additional penalty on the coefficients.

   - C) Increase the magnitude of the coefficients.

   - D) Minimize the L2 norm of the coefficients.

Which of the following is the regularization term used in Lasso regression?

- A) $\sum_{j=1}^{n} w_j^2$

- B) $\sum_{j=1}^{n} |w_j|$

- C) $\sum_{j=1}^{n} w_j^3$

- D) $\sum_{j=1}^{n} w_j$

What is the effect of the L1 regularization term in Lasso regression?

- A) It increases the magnitude of the weights.

- B) It causes some weights to become zero, effectively performing feature selection.

- C) It forces the weights to have the same magnitude.

- D) It penalizes the error term.

How does the Lasso regression help in improving the model's performance when working with high-dimensional data?

- A) By increasing the complexity of the model.

- B) By performing feature selection and shrinking coefficients to zero.

- C) By overfitting the model to the training data.

- D) By using a higher-degree polynomial model.

What happens when λ is set to a very high value in Lasso regression?

- A) The model will overfit the data.

- B) The coefficients will be driven to zero, potentially leading to underfitting.

- C) The model will have high variance and low bias.

- D) The model will perform feature selection without any regularization.

Given the following cost function for Ridge regression:

$$\text{Cost Function} = \text{MSE} + \lambda \sum_{j=1}^{n} w_j^2$$

What happens when λ is set to 0?

- A) The regularization term is removed, and the cost function becomes the plain MSE.

- B) The coefficients are set to zero.

- C) The cost function becomes undefined.

- D) The model is over-regularized.

Given a dataset with many correlated features, which of the following models would typically perform better in terms of model interpretation?

- A) Lasso regression, as it will shrink some coefficients to zero, performing feature selection.
- B) Ridge regression, as it will keep all features but shrink their coefficients uniformly.
- C) Linear regression, as it does not include any regularization.
- D) Both Lasso and Ridge regression will perform equally well.

Suppose you are working with a dataset where some features are irrelevant or redundant. Which regularization technique would you prefer and why?

- A) Lasso, because it can shrink the coefficients of irrelevant features to zero, effectively performing feature selection.
- B) Ridge, because it will keep all features but shrink their coefficients uniformly, handling multicollinearity better.
- C) Both Lasso and Ridge would work equally well in this case.
- D) Neither, because regularization should not be applied to such datasets.

If you are using Ridge regression with a very large value for λ, what would likely happen to your model?

- A) The model would overfit the data.
- B) The model would underfit the data, as the coefficients would be shrunk too much.
- C) The model would have high variance and low bias.
- D) The model would perform feature selection and ignore some features.

For a given dataset with several features, which regularization method would you recommend to perform automatic feature selection?

- A) Lasso regression, because it can shrink some coefficients exactly to zero.

- B) Ridge regression, because it works better for high-dimensional data with many correlated features.

- C) Linear regression, because regularization is not needed.

- D) Both Ridge and Lasso regression should be used interchangeably.

Prove below sentences

Lasso regression has **automatically selected features** by setting the coefficient of X_2 to zero, effectively removing it from the model. This is a key characteristic of Lasso regularization, which can be particularly useful in datasets with many features, as it helps to reduce overfitting by eliminating irrelevant features.

-Lasso regression uses L1 regularization, which adds a penalty based on the absolute values of the coefficients, that's what makes Lasso a powerful tool for feature selectios, especially useful when you have lots of variables and want a simpler, more interpretable model.