Linear Regression Interview Questions & Answers

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

A) TRUE  
B) FALSE

**Solution: (A)**

Yes, Linear regression is a supervised learning algorithm because it uses true labels for training. A supervised machine learning model should have an input variable (x) and an output variable (Y) for each example.

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

A) TRUE  
B) FALSE

**Solution: (A)**

**Linear Regression** has dependent variables that have continuous values.

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

A) TRUE  
B) FALSE

**Solution: (A)**

True. A Neural network which is a component of deep learning, can be used as a *universal* approximator, so it can definitely implement a linear regression algorithm.

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

**Solution: (A)**

In linear regression, we try to minimize the least square errors of the model to identify the line of best fit.

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

**Solution: (D)**

Since linear regression gives output as continuous values, so in such cases, we use mean squared error or r-squared metric to evaluate the model performance. The remaining options are used in case of a classification problem that can be solved by logistic regression or decision trees.

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

A) TRUE  
B) FALSE

**Solution: (A)**

True, In the case of lasso regression, we apply an absolute penalty which makes some of the coefficients zero. Lasso (Least Absolute Shrinkage and Selection Operator) regularization is a technique used in machine learning and statistics to prevent overfitting and to promote feature selection by adding a penalty term to the objective function being optimized.

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

**Solution: (A)**

Residuals refer to the error values of the model. Therefore lower residuals that have normal distribution are desired.

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

Now Imagine that you are applying linear [regression](https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-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

**Solution: (B)**

The absolute value of the correlation coefficient denotes the strength of the relationship. Since the absolute correlation is very high, we infer that the relationship is strong between X1 and Y.

Q9. Looking at the below 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

**Solution: (D)**

We cannot comment on the correlation coefficient by using only statement 1.  We need to consider both of these two statements. Consider V1 as x and V2 as |x|. The correlation coefficient would not be close to 1 in such a case.

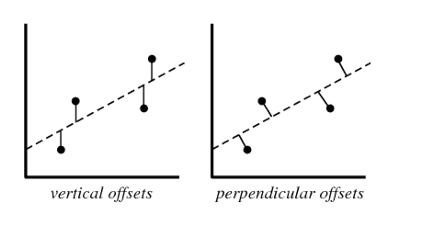
Q10. Suppose the Pearson correlation between V1 and V2 is zero. In such a case, is it right to conclude that V1 and V2 do not have any relation between them?

A) TRUE  
B) FALSE

**Solution: (B)**

Pearson correlation coefficient between 2 variables might be zero even when they have a relationship between them. If the correlation coefficient is zero, it just means that they don’t move together. We can take examples like y=|x| or y=x^2.

Q11. Suppose the horizontal axis is an independent variable and the vertical axis is a dependent variable. Which of the following offsets do we use in linear regression’s least square line fit?



1. Vertical Offsets

B) Perpendicular offset  
C) Both, depending on the situation  
D) None of above

**Solution: (A)**

We always consider residuals as vertical offsets. We calculate the direct differences between the actual value and the Y labels. Perpendicular offsets are useful in the case of dimensionality reduction techniques like PCA.

Q12. True- False: Overfitting is more likely when you have a huge amount of data to train.

A) TRUE  
B) FALSE

**Solution: (B)**

With a small training dataset, it’s easier to find a hypothesis to fit the training data exactly, i.e., overfitting.

Q13. 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

**Solution: (D)**

Instead of gradient descent, a Normal Equation of linear algebra can also be used to find coefficients. Refer to this [article](http://eli.thegreenplace.net/2014/derivation-of-the-normal-equation-for-linear-regression/) to read more about the normal equation.

Q14. 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:**

1. Scale is the same in both graphs for both axes.
2. X-axis is the independent variable, and Y-axis is the dependent variable.

A comparison of blue and black dots

Description automatically generated

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

**Solution: (C)**

The sum of residuals will always be zero; therefore, both have the same sum of residuals.

**Context for Questions 15-17:**

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

Q15. 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

**Solution: (B)**

If the penalty is very large, it means the model is less complex; therefore, the bias would be high.

Q16. 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

**Solution: (B)**

In Lasso, some of the coefficient values become zero, but in the case of Ridge, the coefficients become close to zero but not zero.

Q17. 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

**Solution: (A)**

As already discussed, lasso applies an absolute penalty, so some of the coefficients will become zero.

Q18. 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

**Solution: (A)**

The slope of the regression line will change due to outliers in most cases. So Linear Regression is sensitive to outliers.

Q19. 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

**Solution: (A)**

There should not be any relationship between predicted values and residuals. If there exists any relationship between them, it means that the model has not perfectly captured the information in the data points.

**Context for Questions 20-22:**

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.

Q20. 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

**Solution: (A)**

Since degree 4 will be more complex(overfitting the data) than the degree 3 model, it will again perfectly fit the data. In such a case, the training error will be zero, but the test error may not be zero. Polynomial regression is useful for non-linear data.

Q21. 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

**Solution: (B)**

If a degree 3 polynomial fits the data perfectly, it’s highly likely that a simpler model (degree 2 polynomial) might underfit the data.

Q22. In terms of bias and variance. Which of the following is true when you fit degree 2 polynomial?

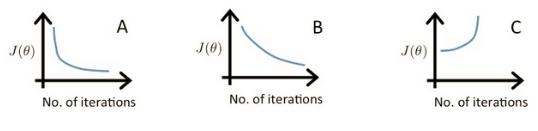
A) Bias will be high, and variance will be high  
B) Bias will be low, and variance will be high  
C) Bias will be high, and variance will be low  
D) Bias will be low, and variance will be low

**Solution: (C)**

Since a degree 2 polynomial will be less complex as compared to degree 3, the bias will be high, and the variance will be low.

**Context for Question 23:**

Below are three graphs, A, B, and C, between the cost function and the number of iterations, I1, I2, and I3, respectively.



Q23. 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

**Solution: (A)**

In the case of a high learning rate, the step will be high, the objective function will decrease quickly initially, but it will not find the global minima, and the objective function starts increasing after a few iterations. In the case of a low learning rate, the step will be small. So the objective function will decrease slowly.

**Context for Questions 24-25:**

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.

Q24. 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

**Solution: (D)**

Training error may increase or decrease depending on the values that are used to fit the model. If the values used to train contain more outliers gradually, then the error might just increase.

Q25. What do you expect will happen with bias and variance as you increase the size of training data?

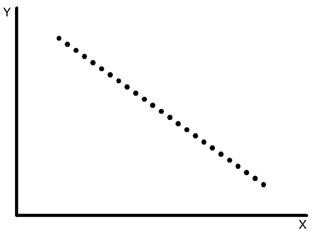
A) Bias increases, and Variance increases  
B) Bias decreases, and Variance increases  
C) Bias decreases, and Variance decreases  
D) Bias increases, and Variance decreases  
E) Can’t Say False

**Solution: (D)**

As we increase the size of the training data, the bias would increase while the variance would decrease.

**Context for Question 26:**

Consider the following data where one input(X) and one output(Y) are given.



Q26. 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

**Solution: (C)**

We can perfectly fit the straight line on the following data so that the mean error will be zero.

**Context for Questions 27-28:**

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 |

Q27. Which of the following scenario would give you the right hyperparameter?

A) 1  
B) 2  
C) 3  
D) 4

**Solution: (B)**

Option B would be the better option because it leads to less training as well as a validation error.

Q28. 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

**Solution: (D)**

If the added feature is important, the training and validation error would decrease.

**Context for Questions 29-30:**

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

Q29. 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

**Solution: (A)**

In case of underfitting, you need to induce more variables in variable space or you can add some polynomial degree variables to make the model more complex to be able to fit the data better.

Q30. 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

**Solution: (D)**

I won’t use any regularization methods because regularization is used in case of overfitting.