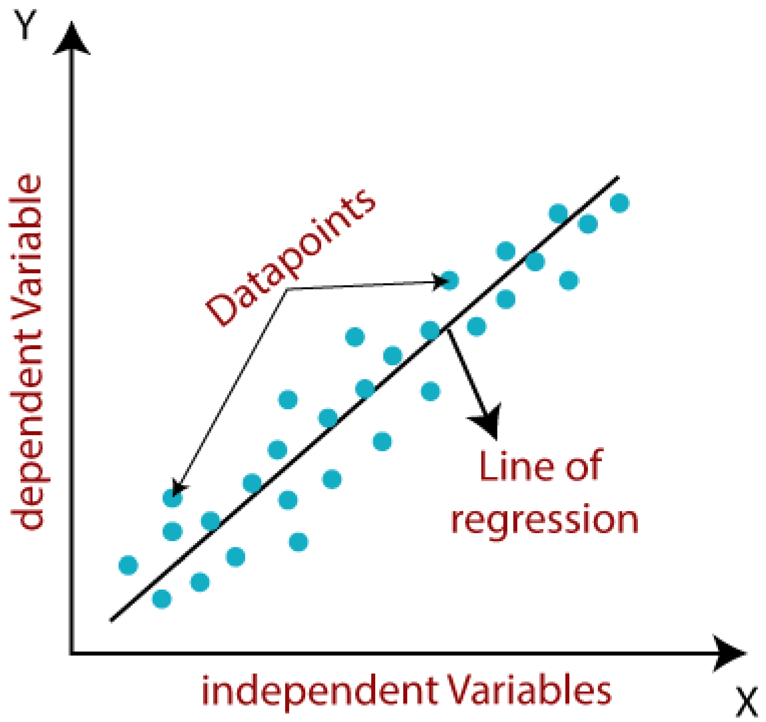
Preparing for interview on Machine Learning? Here, is a complete guide to interview questions on Linear Regression.

Writuparna Banerjee Follow
Aug 8, 2020 · 12 min read

Hey! Are you ready for the interview? Still not confident!!!! Don't worry, go through the commonly asked interview questions on Linear Regression. To let you know, it is a common practice to test data science aspirants on commonly used machine learning algorithms. These most commonly used conventional algorithms being linear regression, logistic regression, decision trees, random forest etc. Data scientists are expected to possess an in-depth knowledge of these algorithms. It is the basis of many different ML Algorithms, so if you make a mistake in giving these answers during an interview, it might be the end of the interview.

Keeping in mind about such young data science aspirants like you, I have covered all the important concepts asked in interview to strengthen your knowledge in at least one of the conventional algorithms.

So, let's get started with Linear Regression!



Interviewer: **What is linear regression?**

Your answer: Linear regression is a method of finding the best straight line fitting to the given data, i.e. finding the best linear relationship between the independent and dependent variables.
In technical terms, linear regression is a machine learning algorithm that finds the best linear-fit relationship on any given data, between independent and dependent variables. It is mostly done by the Sum of Squared Residuals Method.

Interviewer: What are the assumptions made in linear regression model?

Your answer: The important assumptions in linear regression analysis are:

1. There should be a linear and additive relationship between dependent (response) variable and independent (predictor) variable(s). A linear relationship suggests that a change in response Y due to one unit change in X is constant, regardless of the value of X. An additive relationship suggests that the effect of X on Y is independent of other variables.

2. There should be no correlation between the residual (error) terms.
3. The independent variables should not be correlated.
4. The error terms must have constant variance. This phenomenon is known as homoskedasticity.
5. The error terms must be normally distributed.

***Interviewer: What* if these assumptions get violated ?**

Your answer: To understand the outcomes of violating such assumptions we have to dive into the assumptions.

Linear and Additive: If we fit a linear model to a non-linear and non-additive data set, the regression algorithm would fail to capture the trend mathematically, thus resulting in an inefficient model. Also, this will result in erroneous predictions on an unseen data set.

Autocorrelation: Autocorrelation occurs when the residuals are not independent from each other. In other words when the value of y(x+1) is not independent from the value of y(x).The presence of correlation in error terms drastically reduces model's accuracy. This usually occurs in time series models where the next instant is dependent on previous instant. If the error terms are correlated, the estimated standard errors tend to underestimate the true standard error. If this happens, it causes confidence intervals and prediction intervals to be narrower.

> Confidence interval is a range of values so defined that there is a specified probability that the value of a parameter lies within it.
>
> Prediction interval is the range that likely contains the value of the dependent variable for a single new observation when specific values of the independent variables are given. Narrower

prediction interval means that the predicted value of a future observation with the same settings would lie in a narrower range.

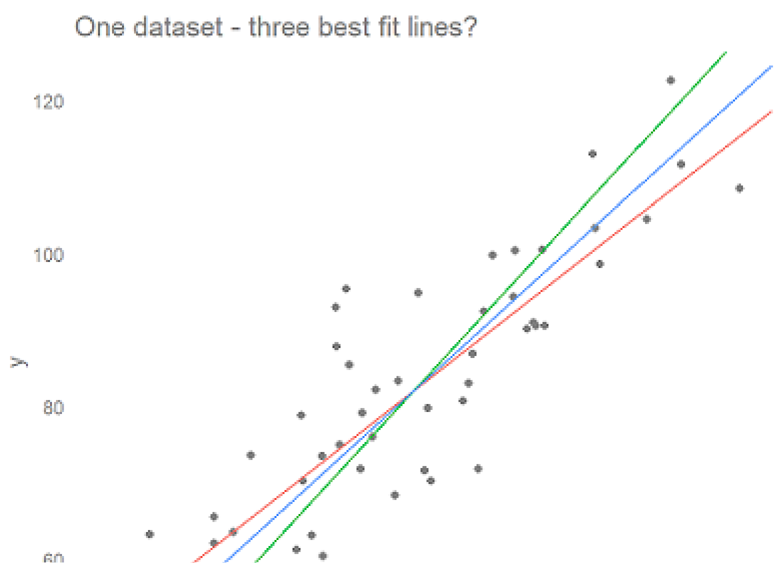
Multicollinearity: This phenomenon exists when the independent variables are found to be moderately or highly correlated. In a model with correlated variables, it becomes a tough task to figure out the true relationship of predictors with response variable. In other words, it becomes difficult to find out which variable is actually contributing to predict the response variable.

Moreover, with presence of correlated predictors, the standard errors tend to increase. And, with large standard errors, the confidence interval becomes wider leading to less precise estimates of slope parameters.

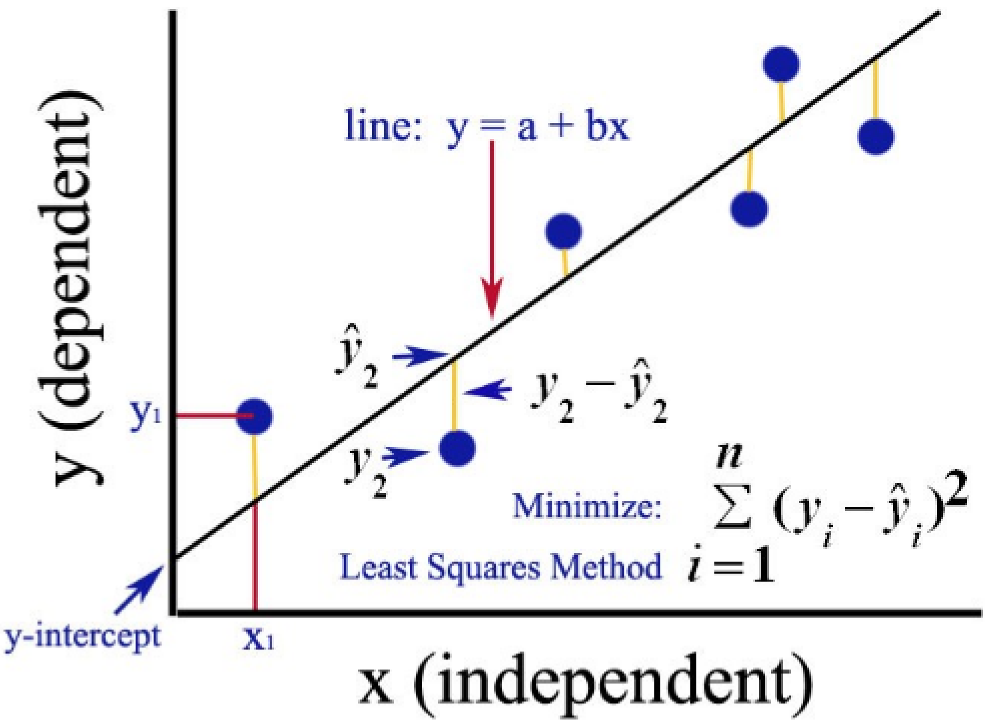
Heteroskedasticity: The presence of non-constant variance in the error terms results in heteroskedasticity. Generally, non-constant variance arises in presence of outliers. It looks like that these values get too much weight, thereby disproportionately influences the model's performance. When this phenomenon occurs, the confidence interval tends to be unrealistically wide or narrow.

Normal Distribution of error terms: If the error terms are non- normally distributed, confidence intervals may become too wide or narrow. Presence of non — normal distribution suggests that there are a few unusual data points which must be studied closely to make a better model.

Interviewer: How to find the best fit line in a linear regression model?



To find the best fit line for our model we have to make the distance with respect to all the points minimum. We have to find that line which is closest to all the points. In statistics, this vertical distance is called residual.



Residual is equal to the difference between the observed value and the predicted value. For data points above the line, the residual is positive, and for data points below the line, the residual is negative. So, if were to find out the sum of all the residuals, then due to the negative errors there will be subtractions in the distance and the the value of resultant distance would be less than the actual. So, to eliminate the negative sign we have to square each residual and find out it's sum.

$$\text{residual} = y - \hat{y}$$
$$\text{sum of squared residuals} = \Sigma(y - \hat{y})^2$$

This distance is known as Sum of Squared Residuals(SSE) and the method is known as Least Squares Method as we need to find that value of m and b of the linear regression line for which SSE is minimum.

Interviewer: **Why do we square the error instead of using modulus?**

It's true that one could choose to use the absolute error instead of the squared error. In fact, the absolute error is often closer to what we want when making predictions from our model. But, we want to penalize those predicted values which is contributing the maximum error. Moreover looking a little deeper, the squared error is everywhere differentiable, while the absolute error is not (its derivative is undefined at 0). This makes the squared error more amenable to the techniques of mathematical optimization. To optimize the squared error, we can just set its derivative equal to 0 and solve. To optimize the absolute error often requires more complex techniques. Actually we find the Root Mean Squared Error so that the unit of RMSE and the dependent variable are equal.

Interviewer: **What are techniques adopted to find the slope and the intercept of the linear regression line which best fits the model?**

Your answer: There are mainly two methods:

1. Ordinary Least Squares(Statistics domain)
2. Gradient Descent(Calculus family)

Interviewer: Explain Ordinary Least Squares Regression in brief.

Your answer: *Ordinary least squares (OLS) regression* is a statistical method of analysis that estimates the relationship between one or more independent variables and a dependent variable. The method estimates the relationship by minimizing the sum of the squares of the difference between the observed and predicted values of the dependent variable configured as a straight line. OLS regression is used in bivariate model, that is, a model in which there is only one independent variable (X) predicting a dependent variable (Y). However, the logic of OLS regression can also be used in multivariate model in which there are two or more independent variables.

Interviewer: **What are the limitations of OLS?**

Your answer: OLS is computationally too expensive. It performs well with small data. For larger data Gradient Descent is preferred.

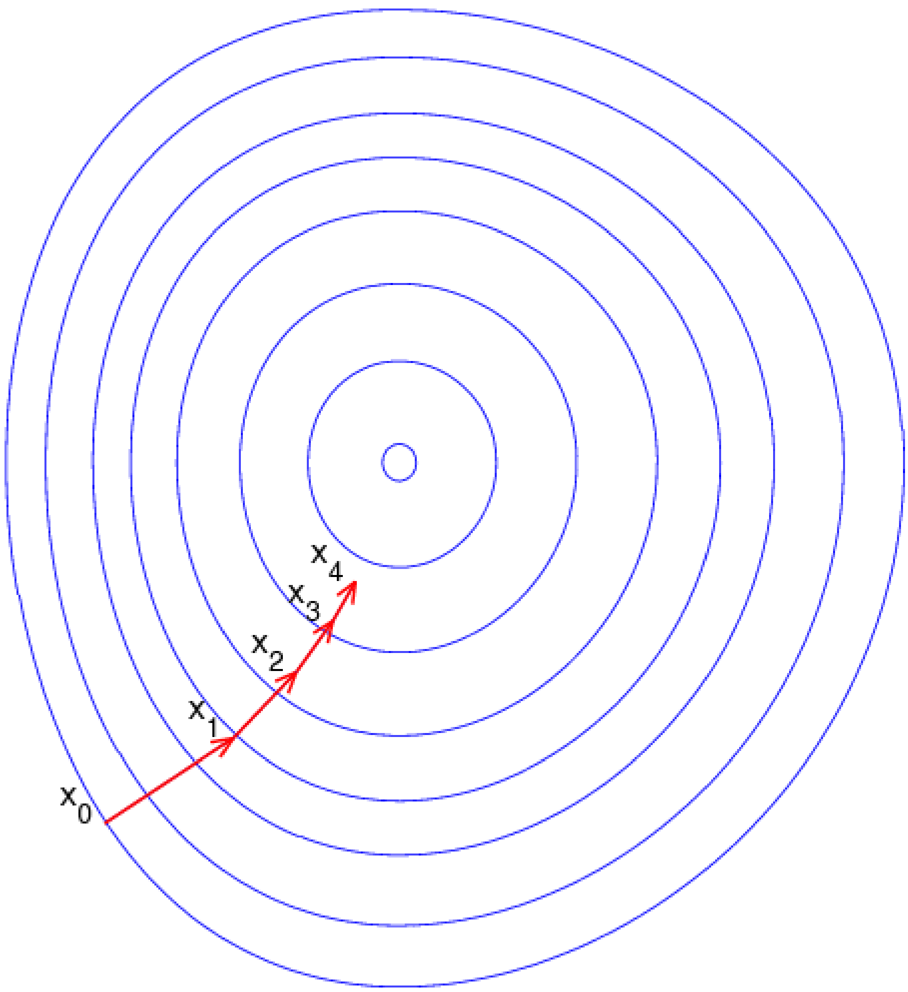
Interviewer: **Can you briefly explain gradient descent?**

Gradient descent is an optimization algorithm that's used when training a machine learning model. It's based on a convex function and tweaks its parameters iteratively to minimize a given function to its local minimum.

We can think of a gradient as the slope of a function. The higher the gradient, the steeper the slope and the faster a model can learn. But if the slope is zero, the model stops learning. In mathematical terms, a gradient is a partial derivative with respect to its inputs.

> Let's imagine a blindfolded man who wants to climb to the top of a hill with the fewest steps along the way as possible. He might start climbing the hill by taking really big steps in the steepest direction, which he can do as long as he is not close to the top. As he comes closer to the top his steps will get smaller and smaller to avoid overshooting it. This process can be described mathematically using the gradient.
>
> Imagine the image below illustrates our hill from a top-down view and the red arrows are the steps

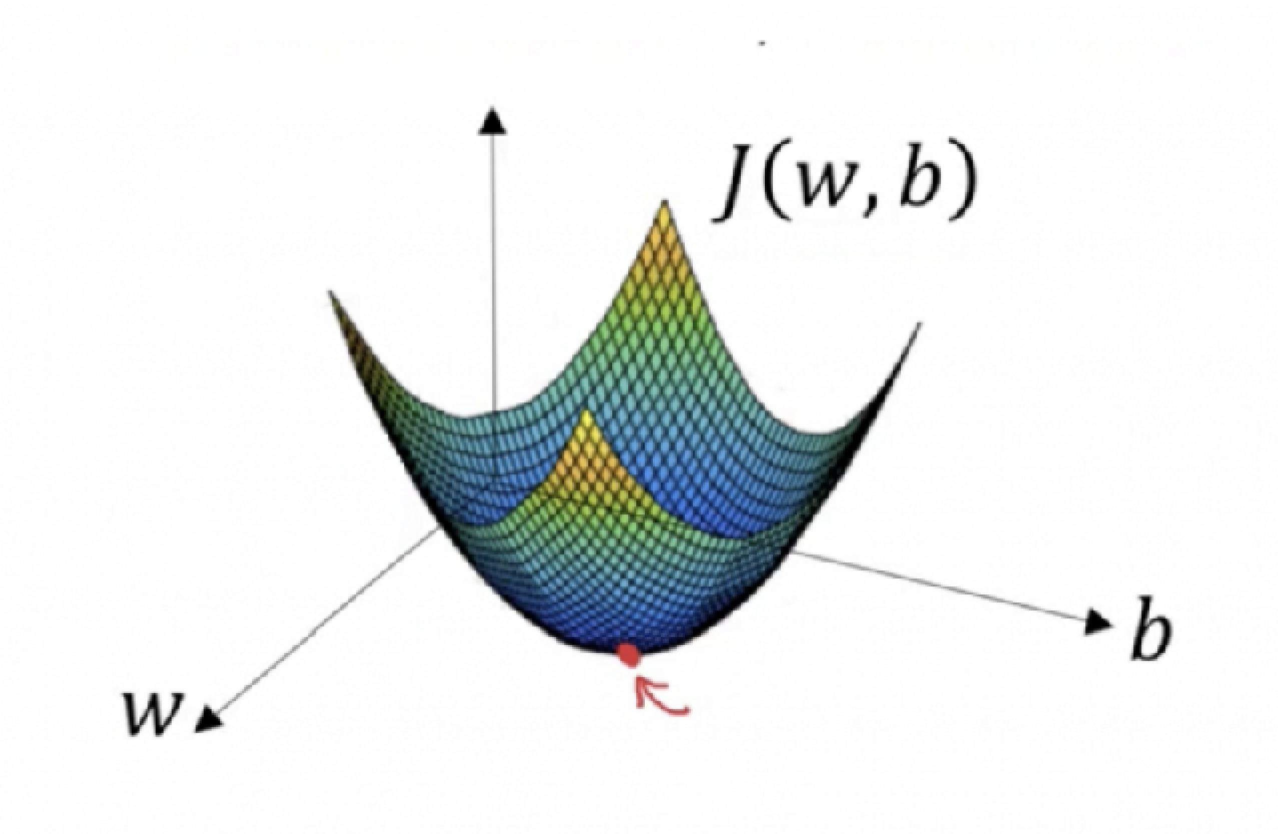
of our climber. Think of a gradient in this context as a vector that contains the direction of the steepest step the blindfolded man can take and also how long that step should be.



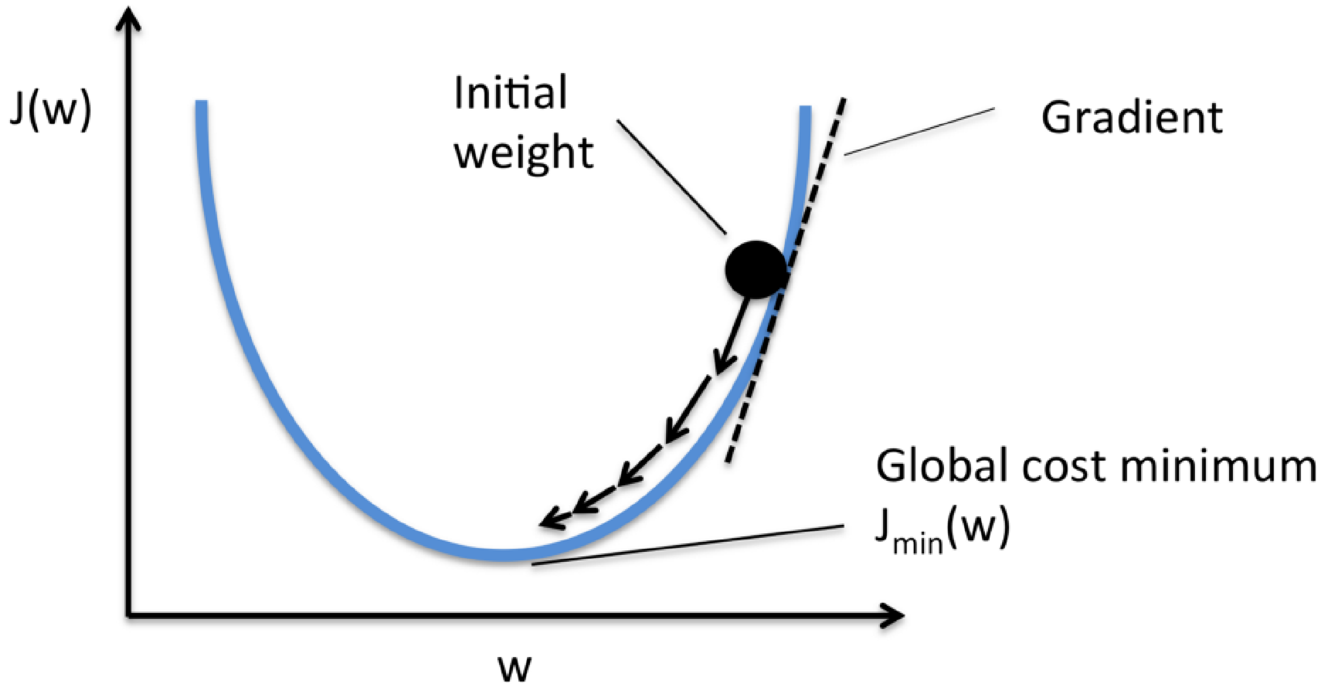
Note that the gradient ranging from X0 to X1 is much longer than the one reaching from X3 to X4. This is because the steepness/slope of the hill, which determines the length of the vector, is less. This perfectly represents the example of the hill because the hill is getting less steep the

> higher it's climbed. Therefore a reduced gradient goes along with a reduced slope and a reduced step size for the hill climber.
>
> Instead of climbing up a hill, think of gradient descent as hiking down to the bottom of a valley. This is a better analogy because it is a minimization algorithm that minimizes a given function.

Let's imagine we have a machine learning problem and want to train our algorithm with gradient descent to minimize our cost-function $J(w, b)$ and reach its local minimum by tweaking its parameters (w and b). We can assume the horizontal axes represent the parameters (w and b), while the cost function $J(w, b)$ is represented on the vertical axes.



We know we want to find the values of w and b that correspond to the minimum of the cost function (marked with the red arrow in the diagram above). To start finding the right values we initialize w and b with some random numbers. Gradient descent then starts at that point (somewhere around the top of our illustration), and it takes one step after another in the steepest downside direction (i.e., from the top to the bottom of the illustration) until it reaches the point where the cost function is as small as possible.



Interviewer: **Explain the significance of learning rate.**

Your answer: Learning rate determines how big are the steps that gradient descent takes into the direction of the local minimum, which figures out how fast or slow we will move towards the optimal weights.

For gradient descent to reach the local minimum we must set the learning rate to an appropriate value, which is neither too low nor too high. This is important because if the steps it takes are too big, it may not reach the local minimum because it bounces back and forth between the convex function of gradient descent. If we set the learning rate to a very small value, gradient descent will eventually reach the local minimum but that may take a while.

Interviewer: **How to evaluate regression models?**

Your answer: There are five metrics used to evaluate regression models:

1. Mean Absolute Error(MAE)
2. Mean Squared Error(MSE)
3. Root Mean Squared Error(RMSE)
4. R-Squared(Coefficient of Determination)
5. Adjusted R-Squared

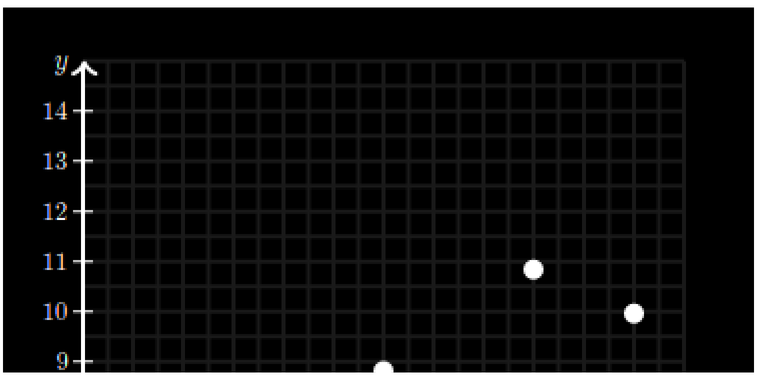
***Interviewer:* Which evaluation technique should you prefer to use for data having a lot of outliers in it?**

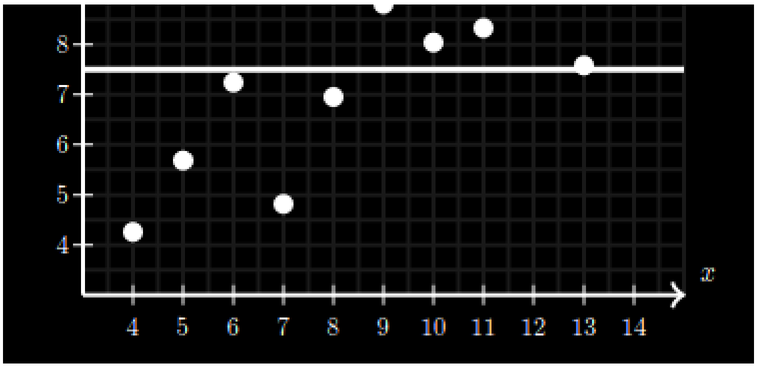
Your answer: Mean Absolute Error(MAE) is preferable to use for data having too many outliers in it because MAE is robust to outliers whereas MSE and RMSE are very susceptible to outliers and starts penalizing the outliers by squaring the residuals.

***Interviewer:* What's the intuition behind R-Squared?**

Your answer: We use linear regression to predict y given some value of x. But suppose that we had to predict a y value without a corresponding x value.

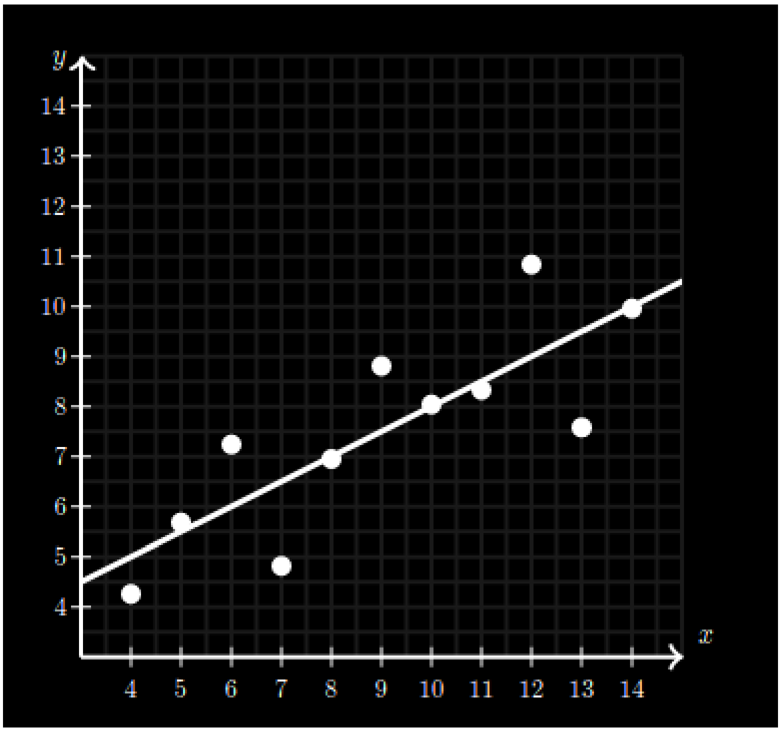
Without using regression on the x variable, our most reasonable estimate would be to simply predict the average of the y values.





However, this line will not fit the data very well(as we can see in the figure above). One way to measure the fit of the line is to calculate the sum of the squared residuals — this gives us an overall sense of how much prediction error a given model has.

Now, if we predict the same data with regression we will see that the least-squares regression line will seem to fit the data pretty well (as shown in the figure below).



We will find that using least-squares regression, the sum of the squared residuals has been considerably reduced.

So using least-squares regression eliminated a considerable amount of prediction error. R-squared tells us what percent of the prediction error in the y variable is eliminated when we use least-squares regression on the x variable.

As a result, R^2 is also called the **coefficient of determination.** Many formal definitions say that R^2 tells us what percent of the variability in the y variable is accounted for by the regression on the x variable. The value of R^2 varies from 0 to 1.

***Interviewer:* Can R^2 be negative?**

Your answer: Yes, R^2 can be negative. The formula of R^2 is given by:

$$R^2 = \frac{Var_{mean} - Var_R}{Var_{mean}}$$

$$= 1 - \frac{Var_R}{Var_{mean}}$$

where,

Var mean = Variance by the mean line

Var R = Variance by the regression line

As variance is equal to the sum of squared error. So, R^2 can be written as:

$$R^2 = 1 - \frac{SS_R}{SS_M} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \overline{y})^2}$$

If the sum of squared error of the mean line(SSM) is greater than the regression line(SSR), R squared will be negative.

***Interviewer:* What are the flaws in R-squared?**

Your answer: There are two major flaws:

Problem 1: R^2 increases with every predictor added to a model. As R^2 always increases and never decreases, it can appear to be a better fit with the more terms we add to the model. This can be completely misleading.

Problem 2: Similarly, if our model has too many terms and too many high-order polynomials we can run into the problem of over-fitting the data. When we over-fit data, a misleadingly high R^2 value can lead to misleading predictions.

***Interviewer:* What is adjusted R^2?**

Your answer: Adjusted R-squared is used to determine how reliable the correlation is between the independent variables and the dependent variable. On addition of highly correlated variables the adjusted R-squared will increase whereas for variables with no correlation with dependent variable the adjusted R-squared will decrease.

The formula is:

$$R^2_{adj} = 1 - \left[\frac{(1-R^2)(n-1)}{n-k-1}\right]$$

where:

- n is the number of points in our data sample.
- k is the number of independent regressors, i.e. the number of input columns.

Adjusted R^2 will always be less than or equal to R^2.

Conclusion: These are the few basic questions that can be asked from Linear Regression. Please note, that I have explained few questions in detail for clearing your concepts and for your better understanding. While answering to the interviewer be specific. Answer only what you have been asked for. Be confident and answer smartly. Good luck!

If you find this helpful, don't forget to hit 👏 icon . *It will help other young aspirants like you to see the story. Thank you!*😊

References:

1. https://builtin.com/data-science/gradient-descent

2. https://www.khanacademy.org/math/ap-statistics/bivariate-data-ap/assessing-fit-least-squares-regression/a/r-squared-intuition

Machine Learning Linear Regression Regression Metric Ordinary Least Square

Gradient Descent

Most Commonly Asked Interview Questions On Linear Regression

Analyzing data is vital in today's era of computers. There is tremendous scope for data scientists and data analysis in the industry today. The companies recruiting these data scientists would naturally interview them to understand their capability. One of the favorite topics on which the interviewers ask questions is 'Linear Regression.' Here are some of the common Linear Regression Interview Questions that pop up in interviews all over the world.

Linear Regression Interview Questions – Fundamental Questions

Table of Contents [hide]

will you require to estimate in a simple linear regression model?
Conclusion

Let us begin with a fundamental Linear Regression Interview Questions

1. What is a Linear Regression?

In simple terms, linear regression is adopting a linear approach to modeling the relationship between a dependent variable (scalar response) and one or more independent variables (explanatory variables). In case you have one explanatory variable, you call it a simple linear regression. In case you have more than one independent variable, you refer to the process as multiple linear regressions.

2. Can you list out the critical assumptions of linear regression?

There are three crucial assumptions one has to make in linear regression. They are,

01. It is imperative to have a linear relationship between the dependent and independent A scatter plot can prove handy to check out this fact.
02. The independent variables in the dataset should not exhibit any multi-collinearity. In case they do, it should be at the barest minimum. There should be a restriction on their value depending on the domain requirement.
03. Homoscedasticity is one of the most critical It states that there should be an equal distribution of errors.

3. What is Heteroscedasticity?

Heteroscedasticity is the exact opposite of homoscedasticity. It entails that there is no equal distribution of the error terms. You use a log function to rectify this phenomenon.

4. What is the primary difference between R square and adjusted R square?

In linear regression, you use both these values for model validation. However, there is a clear distinction between the two. R square accounts for the variation of all independent variables on the dependent variable. In other words, it considers each independent variable for explaining the variation. In the case of Adjusted R square, it accounts for the significant variables alone for indicating the percentage of variation in the model. By significant, we refer to the P values less than 0.05.

5. Can you list out the formulas to find RMSE and MSE?

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Predicted_i - Actual_i)^2}{N}}$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2$$

The most common measures of accuracy for any linear regression are RMSE and MSE. MSE stands for Mean Square Error whereas RMSE stands for Root Mean Square Error. The formulas of RMSE and MSE are as hereunder

Mean Square Error. The formulas of RMSE and MSE are as hereunder.

Linear Regression Interview Questions – Complex Questions

We have seen some of the basic interview questions on linear regression. As we move further into the article, we shall look at some of the complex linear regression interview questions as well.

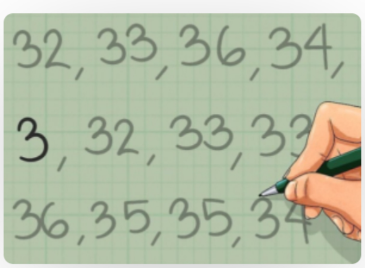
6. Can you name a possible method of improving the accuracy of a linear regression model?

You can do so in many ways. One of the most common ways is 'The Outlier Treatment.'

Outliers have great significance in linear regression because regression is very sensitive to outliers. Therefore, it becomes critical to treat outliers with appropriate values. It can also prove useful if you replace the values with mean, median, mode or percentile depending on the distribution.

7. What are outliers? How do you detect and treat them?

An outlier is an observation point distant from other observations. It might be due to a variance in the measurement. It can also indicate an experimental error. Under such circumstances, you need to exclude the same from the data set. If you do not detect and treat them, they can cause problems in statistical analysis.



You can see that 3 is the outlier in this example.

There is no strict mathematical calculation of how to determine an outlier. Deciding whether an observation is an outlier or not, is itself a subjective exercise. However, you can detect outliers through various methods. Some of them are graphical and are known as normal probability plots whereas some are model-based. You have some hybrid techniques such as Boxplots.

Once you have detected the outlier, you should either remove them or correct them to ensure accurate analysis. Some of the methods of eliminating outliers are the Z-Score and the IQR Score methods.

8. How do you interpret a Q-Q plot in a linear regression model?

As the name suggests, the Q-Q plot is a graphical plotting of the quantiles of two distributions with respect to each other. In other words, you plot quantiles against quantiles.

Whenever you interpret a Q-Q plot, you should concentrate on the 'y = x' line. You also call it the 45-degree line in statistics. It entails that each of your distributions has the same quantiles. In case you witness a deviation from this line, one of the distributions could be skewed when compared to the other.

9. What is the importance of the F-test in a linear model?

The F-test is a crucial one in the sense that it tests the goodness of the model. When you reiterate the model to improve the accuracy with the changes, the F-test proves its utility in understanding the effect of the overall regression.

10. What are the disadvantages of the linear regression model?

One of the most significant demerits of the linear model is that it is sensitive and dependent on the outliers. It can affect the overall result. Another notable demerit of the linear model is overfitting. Similarly, underfitting is also a significant disadvantage of the linear model.

11. What is the curse of dimensionality? Can you give an example?

When you analyze and organize data in high-dimensional spaces (usually in thousands), various situations can arise that usually do not do so when you analyze data in low-dimensional settings (3-dimensional physical space). The curse of dimensionality refers to such phenomena.

Here is an example.

All kids love to eat chocolates. Now, you bring a truckload of chocolates in front of the kid. These chocolates come in different colors, shapes, tastes, and price. Consider the following scenario.

The kid has to choose one chocolate from the truck depending on the following factors.

01. Only taste – There are usually four tastes, sweet, salty, sour, and bitter. Hence, the child will have to try out only four chocolates before choosing one to its liking.
02. Taste and Color – Assume there are only four colors. Hence, the child will now have to taste a minimum of 16 (4 X 4) before making the right choice.
03. Taste, color, and shape – Let us assume that there are five shapes. Therefore, the child will now have to eat a minimum of 80 chocolates (4 X 4 X 5).

What will happen to the child if it tries out 80 chocolates at a time? It will naturally become sick. Hence, it will not be in a position to try out the chocolates. This example is the perfect one to explain the curse of dimensionality. The more the options you have, the more the problems you encounter.

Linear Regression Interview Questions – Multiple Choice Questions

Let us now look at some multiple-choice linear regression interview questions.

1. In regression analysis, which of the statements is true?

01. The mean of residuals is always equal to Zero
02. The Mean of residuals is less than Zero at all times
03. The Mean of residuals is more than Zero at all times
04. You do not have any such rule for residuals.

The correct answer is A. In regression analysis, the sum of the residuals in regression is always equal to Zero. Thus, it implies that the mean will also be Zero if the sum of the residuals is Zero.

2. Which of the statements is correct about Heteroscedasticity?

01. Linear regression with different error terms
02. Linear regression with constant error terms
03. Linear regression with no error terms
04. None of the above

The solution is the option A. When you have a non-constant variance in the error terms, it results in Heteroscedasticity. Such non-constant variance occurs because you have outliers.

3. Which of the following plots is best suited to test the linear relationship of independent and dependent continuous variables?

01. Scatter Plot
02. Bar Chart
03. Histograms
04. None of the above options

The answer is A. The Scatter plot is the best way to determine the relationship between continuous variables. You can find out how one variable changes with respect to the other.

4. If you have only one independent variable, how many coefficients will you require to estimate in a simple linear regression model?

01. One
02. Two
03. No idea

The answer is B. Consider the simple linear regression with one independent variable. Y = a + bx. You can see that you need two coefficients.

Conclusion

We segregated the Linear Regression Interview Questions on the basis of fundamental, complex and multiple-choice questions.

Hopefully, you must have understood the above mentioned Linear Regression Interview Questions thoroughly.

We have seen the importance of linear regression through some interview questions on linear regression. Data scientists need to master this aspect as linear regression is usually a favorite topic with the interviewers. You can refer to the official website of Digital Vidya to learn more about the important aspects of Data Science.

Srinivasan

Srinivasan, more popularly known as Srini, is the person to turn to for writing blogs and informative articles on various subjects like banking, insurance, social media marketing, education, and product review descriptions. Writing articles on digital marketing and social media marketing comes naturally to him. Similarly, he has the capacity and more importantly, the patience to do in-depth research before committing anything on paper.

← Prev: Search Engine Penalties in 2019: Top 10 Algorithm Updates

Next: A Beginner's Guide to Credit Risk Modelling →

2 Comments

Kumarjoshua on October 26, 2018 at 17:03

Nice post, great content. Thank you for sharing.

REPLY

Digital Vidya Editorial Team on December 17, 2018 at 12:50

Thanks for the appreciation.

REPLY

Submit a Comment

Your email address will not be published. Required fields are marked *

Comment

Name *

Email *

☐ Save my name, email, and website in this browser for the next time I comment.

SUBMIT COMMENT

Discuss With A Career Advisor

Not Sure, What to learn and how it will help you?

Schedule a Call