

# MACHINE LEARNING INTERVIEW QUESTIONS ( Part-1 )

Submitted by :- Ambarish Singh

## Q.1) What is Linear Regression ?

### Ans.1 :-

Linear regression is a regression model that estimates the relationship between one independent variable and one dependent variable using a straight line. Both variables should be quantitative.

For example, the relationship between Number of Rooms and Price of the House can be modeled using a straight line: as Number of Rooms increases, the Price of House also increases. This linear relationship is so certain that we can tell that Price of the House can Predicts by Number of Rooms.

**Three Major uses for Regression Analysis are :-**

- Determining the Strength of Predictors.
- Forecasting an Effect
- Trend Forecasting.

## Q.2) How we can Calculate Error in Linear Regression ?

### Ans.2 :-

Linear regression most often uses mean-square error (MSE) to calculate the error of the model.

**MSE is calculated by:**

- Measuring the distance of the observed y-values from the predicted y-values at each value of x;
- Squaring each of these distances;
- Calculating the mean of each of the squared distances.

Linear regression fits a line to the data by finding the regression coefficient that results in the smallest MSE.

## Q.3) Different Between Loss and Cost function ?

### Ans.3 :-

**Loss Function :-** A loss function is a function which measures the error between a single prediction and the corresponding actual value.

**Cost Function :-** A cost function is a function which measures the error between predictions and their actual values across the whole dataset.

Both the cost and the loss function calculate prediction error, but the key difference is the level at which they calculate this error. A loss function calculates the error per observation, whilst the cost function calculates the error for all observations by aggregating the loss values.

### Q.4) What is MAE, MSE and RMSE ?

#### Ans.4 :-

**MAE :-** The Mean absolute error (MAE) represents the average of the absolute difference between the actual and predicted values in the dataset. It measures the average of the residuals in the dataset.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}|$$

Where,

$\hat{y}$  – predicted value of  $y$

$\bar{y}$  – mean value of  $y$

**MSE :-** Mean Squared Error (MSE) represents the average of the squared difference between the original and predicted values in the data set. It measures the variance of the residuals.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2$$

**RMSE :-** Root Mean Squared Error (RMSE) is the square root of Mean Squared error. It measures the standard deviation of residuals.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2}$$

### **Q.5) Explain How Gradient Descent works in Linear Regression ?**

#### **Ans.5 :-**

Gradient descent is an optimization algorithm used to minimize some function by iteratively moving in the direction of steepest descent as defined by the negative of the gradient. In machine learning, we use gradient descent to update the parameters of our model.

When there are more than one inputs you can use a process of optimizing values of coefficients by iteratively minimizing error of model on your training data. This is called Gradient Descent and works by starting with random values for each coefficient. The sum of squared errors are calculated for each pair of input and output variables.

A learning rate is used for each pair of input and output values. It is a scalar factor and coefficients are updated in direction towards minimizing error. The process is repeated until a minimum sum squared error is achieved or no further improvement is possible.

When using this method, learning rate alpha determines the size of improvement step to take on each iteration of procedure. In practise, Gradient Descent is useful when there is a large dataset either in number of rows or number of columns.

In short ,it is a minimization algorithm meant for minimizing a given activation function.

- To understand this in simple terms, let us see an example on how it works...



### Example of Gradient Descent Algorithm

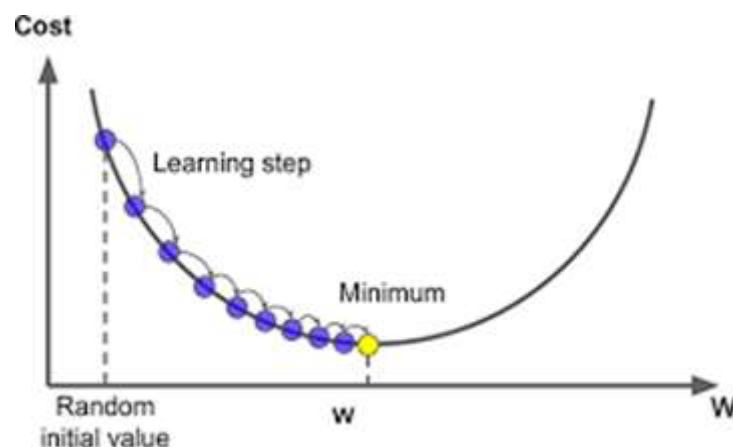
Imagine a valley and a person with no sense of direction who wants to get to the bottom of the valley. He goes down the slope and takes large steps when the slope is steep and small steps when the slope is less steep. He decides his next position based on his current position and stops when he gets to the bottom of the valley which was his goal.

**Now,** Coming to our topic, The Gradient Descent algorithm determines the values of  $m$  and  $c$ , such that line corresponding to those values is best fitting line / gives minimum error.

First we need to calculate the loss function. The loss function is defined as the error in our predicted value of  $m$  and  $c$ . Our aim is to minimize this error to obtain most accurate value of  $m$  and  $c$ . To calculate the loss, Mean Squared Error function is used.

- The difference between actual and predicted  $y$  value is found.
- The difference is squared.
- Then, the mean of squares for every value in  $x$  is calculated.

Now, we have to minimize " $m$ " and " $c$ ". To minimize these parameters, Gradient Descent Algorithm is used.



## process of gradient descent algorithm

1. Initially, let  $m = 0$ ,  $c = 0$

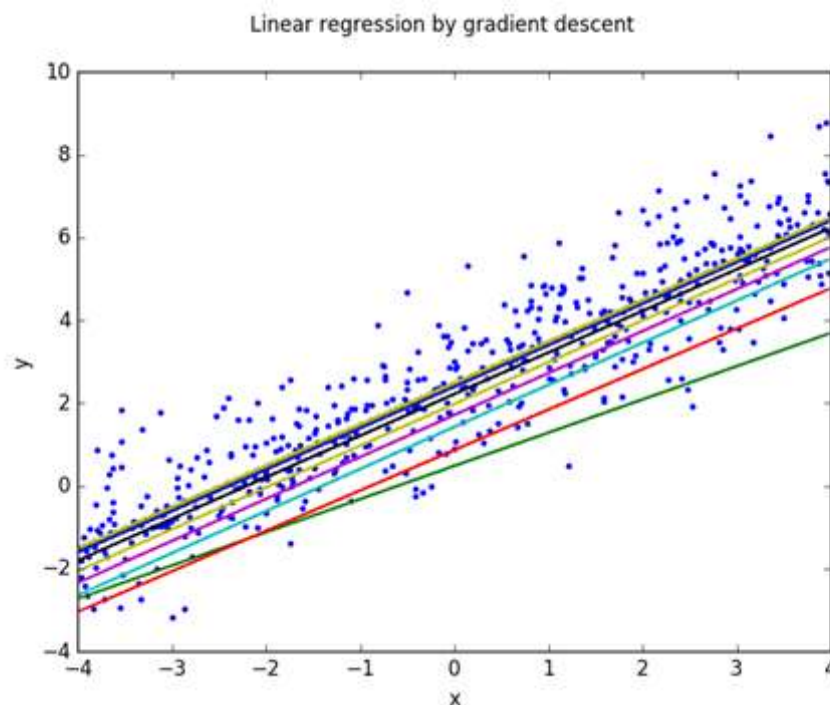
Where  $L$  = learning rate — controlling how much the value of “ $m$ ” changes with each step. The smaller the  $L$ , greater the accuracy.  $L = 0.001$  for a good accuracy.

2. Calculating the partial derivative of loss function wrt “ $m$ ” and give current values of  $x$ ,  $y$ ,  $m$  and  $c$  to get the derivative  $D$ .

3. Similarly,  $D$  wrt  $c$

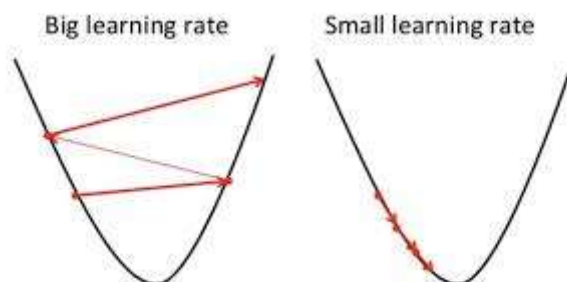
4. Now, updating the current value of  $m$  and  $c$ ,

5. We repeat this process until loss function is very small ,i.e. ideally 0 % error (100% accuracy).



- every line depicts each iteration in the process of arriving at the best line.

**Going back to analogy**,  $m$  = current position,  $D$  = steepness of slope,  $L$  = speed at which person moves (learning rate). Every new “ $m$ ” value calculated is the next position.



When the learning rate is quite high, the process will be a very haphazard one. So, the smaller the

learning rate, the better the result of the model will be.

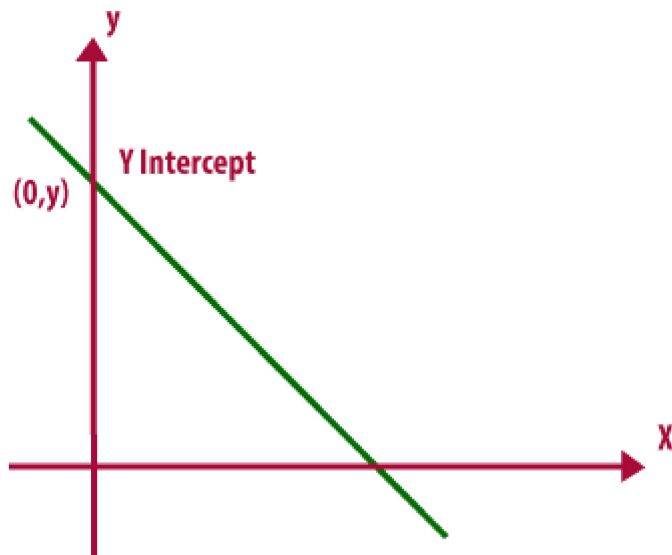
After repeating this process for many times, the person arrives at the bottom of the valley. With optimum value of  $m$  and  $c$ , model is ready.

**NOTE:** In the case of linear regression model, there is only one minimum and it is the global minimum. Hence we say that the required coefficients for the line is found.

## Q.6) Explain what the Intercept term means ?

### Ans.6 :-

The Intercept (Often labeled as Constant) is the point where the function crosses the  $y$ - Axis. In some Analysis , the regression model only becomes Significant when we remove the intercept, and the regression line reduces to  $Y = bX + \text{Error}$ .



## Q.7) Write all the Assumptions of Linear Regression ?

### Ans.7 :-

Linear regression is an Analysis that assesses whether one or more Predictor variables explain the dependent (Criteria) variable.

**The Regression has Five Key Assumptions:-**

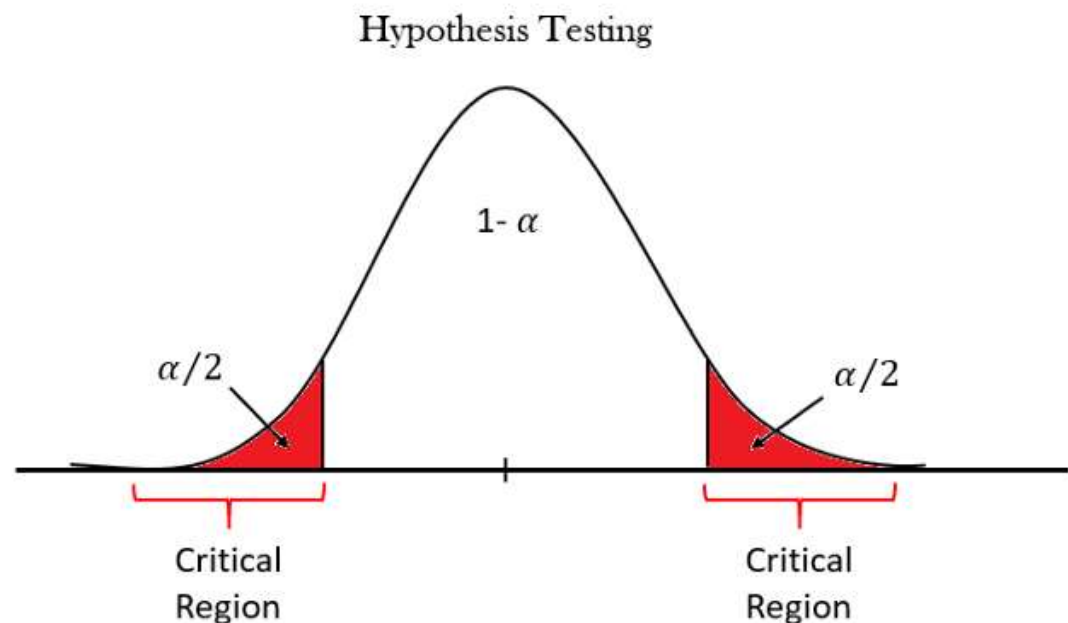
1. The Two Variables Should be in a Linear Relationship.

2. All the Variables Should be Multivariate Normal.
3. There Should be No Multicollinearity in the Data.
4. There Should be No Autocorrelation in the Data.
5. There Should be Homoscedasticity Among the Data.

## Q.8) How is Hypothesis Testing using in Linear Regression ?

### Ans.8 :-

Hypothesis Testing is used to confirm if our beta coefficients are significant in a linear Regression model. Every Time we run the linear Regression Model, we test if the line is significant or not by checking if the coefficient is significant.



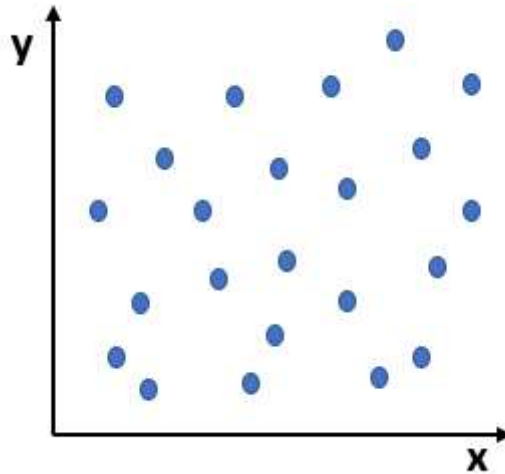
Let us understand it in Simple Linear Regression first.

When we fit a straight line through the data, we get two parameters i.e., the intercept ( $\beta_0$ ) and the slope ( $\beta_1$ ).

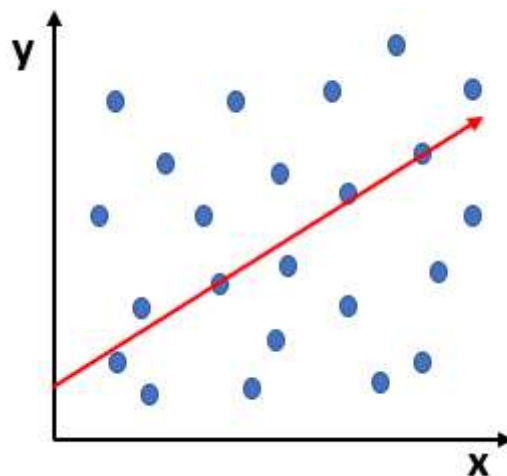
$$y = \beta_0 + \beta_1 x$$

Now,  $\beta_0$  is not of much importance right now, but there are a few aspects around  $\beta_1$  which needs to be checked and verified. Suppose we have a dataset for which the scatter plot looks like the following:

Scatter Plot :-



When we run a linear regression on this dataset in Python, Python will fit a line on the data which looks like the following:



We can clearly see that the data is randomly scattered and doesn't seem to follow a linear trend. Python will anyway fit a line through the data using the least squares method. We can see that the fitted line is of no use in this case. Hence, every time we perform linear regression, we need to test whether the fitted line is a significant one or not (in other terms, test whether  $\beta_1$  is significant or not). We will use Hypothesis Testing on  $\beta_1$  for the same.

#### **Steps to Perform Hypothesis testing:-**

1. Set the Hypothesis.
2. Set the Significance Level, Criteria for a decision.
3. Compute the test statistics.



4. Make a decision.

### **Q.9) How would you decide the Importance of variable for Multivariate Regression ?**

#### **Ans.9 :-**

1. Variables that are already proven in the literature to be related to the outcome.
2. Variables that can either be considered the cause of the exposure, the outcome or both.
3. Interaction terms of Variables that have large main effects.

### **Q.10) What is the Different between R-Squared vs Adjusted R-Squared ?**

#### **Ans.10 :-**

Adjusted R-Squared can be calculated mathematically in terms of sum of Squares. The only Difference between R-Squared and Adjusted R-Squared is Degree of Freedom..

Adjusted R-Squared value can be calculated based on value of R-squared, Number of Independent variables (Predictors), Total Sample size.

**THANK YOU**