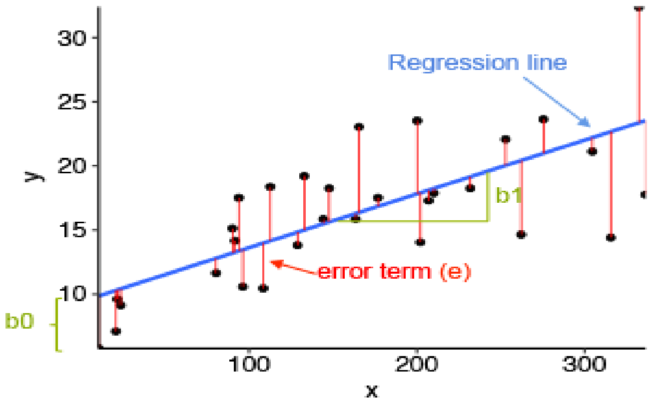
MACHINE LEARNING INTERVIEW QUESTIONS

(1). What is linear regression?

In the Most simple word, We can say Linear Regression is the supervised machine learning algorithm in which the model finds the best fit line between the independent and dependent variable. i.e it finds the linear relationship between the continuous dependent variable and independent variable

A Linear Regression model's main aim is to find the best fit linear line and the optimal values of intercept and coefficients such that the error is minimized.



Types of Linear Regression

1. **Simple Linear Regression**
   Simple Linear Regression is where only one independent variable is present and the model has to find the linear relationship of it with the dependent variable

2. **Multiple Linear Regression**
   Multiple Linear Regression there is more than one independent variable for the model to find the relationship.

   Equation of Simple Linear Regression, where b_o is the intercept, b_1 is the coefficient or slope, x is the independent variable and y is the dependent variable.

$$y = b_o + b_1 x$$

Equation of Multiple Linear Regression, where bo is the intercept, $b_1, b_2, b_3, b_4 \ldots, b_n$ are coefficients or slopes of the independent variables $x_1, x_2, x_3, x_4 \ldots, x_n$, and y is the dependent variable.

$$y = b_o + b_1 x_1 + b_2 x_2 + b_3 x_3 \ldots . + b_n x_n$$

(2). Difference between loss and cost function

LOSS FUNCTION:- **The loss function quantifies how much a model's prediction deviates from the ground truth for one particular object.** So, when we calculate loss, we do it for a single object in the training or test sets.

the Square and Absolute Losses in Regression

Very often, we use the **square(d) error** as the loss function in regression problems:

$$L_{square}(\hat{y}, y) = (\hat{y} - y)^2$$

Another loss function we often use for regression is the **absolute loss**:

$$L_{abs}(\hat{y}, y) = |\hat{y} - y|$$

COST FUNCTION:- Cost function measures the performance of a machine learning model for a data set. The cost function quantifies the error between predicted and expected values and presents that error in the form of a single real number. Depending on the problem, the cost function can be formed in many different ways. The purpose of the cost function is to be either minimized or maximized. For algorithms relying on gradient descent to optimize model parameters, every function has to be differentiable.

the mean Square and mean Absolute error in Regression

Mean Square Error

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

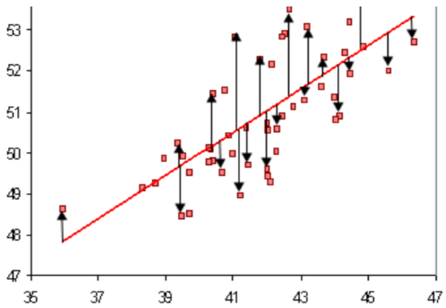
Mean absolute error

$$MAE = \frac{1}{m} \sum_{i=1}^{m} |\hat{y}^{(i)} - y^{(i)}|$$

(3). How can you calculate error in linear regression?

An **error term** in statistics is a value that represents how observed data differs from actual data. It can also be a variable that represents how a given statistical model differs from reality. The error term is often written ε.

The difference between predicted data points and actual data points is known as error/residuals.



We Can Calculate the error in Linear Regression by various methods. Some are the Followings:-

1. Mean Square Error(MSE)

2. Mean Absolute Error(MAE)

3. Root Mean Square Error(RMSE)

(4). What are MAE MSE and RMSE?

1. Mean Square Error

Mean squared error is one of the most commonly used regression metrics. MSE represents the average squared difference between the predictions and expected results. In other words, MSE is an alteration of MAE where, instead of taking the absolute value of differences, we square those differences.

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

2. Mean Absolute Error

Mean absolute error is a regression metric that measures the average magnitude of errors in a group of predictions, without considering their directions. In other words, it's a mean of absolute differences among predictions and expected results where all individual deviations have even importance.

$$MAE = \frac{1}{m} \sum_{i=1}^{m} |\hat{y}^{(i)} - y^{(i)}|$$

3. Root Mean Square Error

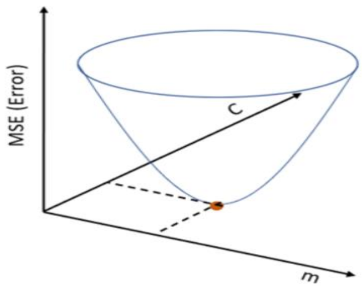
Root Mean Squared Error is the square root of Mean Squared error. It measures the standard deviation of residuals.

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

(5). Explain how gradient descent works in linear regression.

Gradient Descent is an algorithm that finds the best-fit line for a given training dataset in a smaller number of iterations.

If we plot m and c against MSE, it will acquire a bowl shape (As shown in the diagram below)



For some combination of m and c, we will get the least Error (MSE). That combination of m and c will give us our best fit line.

The algorithm starts with some value of m and c (usually starts with m=0, c=0). We calculate MSE (cost) at point m=0, c=0. Let say the MSE (cost) at m=0, c=0 is 100. Then we reduce the value of m and c by some amount (Learning Step). We will notice a decrease in MSE (cost). We will continue doing the same until our loss function is a very small value or ideally 0 (which means 0 error or 100% accuracy).

Step-by-Step Algorithm:

1. Let m = 0 and c = 0. Let L be our learning rate. It could be a small value like 0.01 for good accuracy.

The learning rate gives the rate of speed where the gradient moves during gradient descent. Setting it too high would make your path unstable, and too low would make convergence slow. Put it to zero means your model isn't learning anything from the gradients.

2. Calculate the partial derivative of the Cost function with respect to m. Let the partial derivative of the Cost function with respect to m be D_m (With little change in m how much the Cost function changes).

$$D_m = \frac{\partial(Cost\ Function)}{\partial m} = \frac{\partial}{\partial m}\left(\frac{1}{n}\sum_{i=0}^{n}(y_i - y_{i\,pred})^2\right)$$

$$= \frac{1}{n}\frac{\partial}{\partial m}\left(\sum_{i=0}^{n}(y_i - (mx_i + c))^2\right)$$

$$= \frac{1}{n}\frac{\partial}{\partial m}\left(\sum_{i=0}^{n}(y_i^2 + m^2x_i^2 + c^2 + 2mx_ic - 2y_imx_i - 2y_ic)\right)$$

$$= \frac{-2}{n}\sum_{i=0}^{n}x_i(y_i - (mx_i + c))$$

$$= \frac{-2}{n}\sum_{i=0}^{n}x_i(y_i - y_{i\,pred})$$

Similarly, let's find the partial derivative with respect to c. Let the partial derivative of the Cost function with respect to c be D_c (With little change in c how much the Cost function changes).

$$D_c = \frac{\partial(Cost\ Function)}{\partial c} = \frac{\partial}{\partial c}\left(\frac{1}{n}\sum_{i=0}^{n}(y_i - y_{i\,pred})^2\right)$$

$$= \frac{1}{n}\frac{\partial}{\partial c}\left(\sum_{i=0}^{n}(y_i - (mx_i + c))^2\right)$$

$$= \frac{1}{n}\frac{\partial}{\partial c}\left(\sum_{i=0}^{n}(y_i^2 + m^2x_i^2 + c^2 + 2mx_ic - 2y_imx_i - 2y_ic)\right)$$

$$= \frac{-2}{n}\sum_{i=0}^{n}(y_i - (mx_i + c))$$

$$\frac{-2}{n}\sum_{i=0}^{n}(y_i - y_{i\,pred})$$

3. Now update the current values of m and c using the following equation:

$$m = m - LD_m$$

$$c = c - LDc$$

4. We will repeat this process until our Cost function is very small (ideally 0).

The gradient Descent Algorithm gives optimum values of m and c of the linear regression equation. With these values of m and c, we will get the equation of the best-fit line and be ready to make predictions.

(6) Explain what the intercept term means?

Intercept is something at the point where the regression line intercepts or crosses the Y-Axis. The **intercept** (sometimes called the “constant”) in a regression model represents the mean value of the response variable when all of the predictor variables in the model are equal to zero. $\hat{y} = \beta_0 + \beta_1(x)$

β_0: The mean value of the response variable when x = 0

(7) Write all assumptions for the linear regression

There are four assumptions associated with a linear regression model:

1. **Linearity**: The relationship between X and the mean of Y is linear.
2. **Homoscedasticity**: The variance of residual is the same for any value of X.
3. **Independence**: Observations are independent of each other.
4. **Normality**: For any fixed value of X, Y is normally distributed.

(8) How is hypothesis testing used in linear regression?

Hypothesis testing is used to confirm if beta coefficients are significant in a linear regression model. While running the model each time needs to check whether the line is significant or not by checking whether the coefficient is significant

Hypothesis testing can be carried out in linear regression for the following purposes:

To check whether a predictor is significant for the prediction of the target variable. Two common methods for this are — By the use of p-values: If the p-value of a variable is greater than a certain limit (usually 0.05), the variable is insignificant in the prediction of the target variable. By checking the values of the regression coefficient: If the value of the regression coefficient corresponding to a predictor is zero, that variable is insignificant in the prediction of the target variable and has no linear relationship with it. To check whether the calculated regression coefficients are good estimators of the actual coefficients. The Null and Alternate Hypotheses used in the case of linear regression, respectively, are:

$\beta 1=0$

$\beta 1\neq 0$

Thus, if we reject the Null hypothesis, we can say that the coefficient $\beta 1$ is not equal to zero and hence, is significant for the model. On the other hand, if we fail to reject the Null hypothesis, it is concluded that the coefficient is insignificant and should be dropped from the model.

(9) How would you decide the importance of variables for the multivariate regression?

Let's understand **Variable Importance** first:

- The Variable feature that impacts the dependent/response variable the most can be termed as the most important variable.
- Goal here is to find the rank ordering of features as per importance

When we fit a multiple regression model, we use the *p*-value in the ANOVA table to determine whether the model, as a whole, is significant. A natural next question to ask is which features, among a larger set of all potential features, are important. We could use the individual *p*-values and refit the model with only significant terms. But, remember that the *p*-values are adjusted for the other terms in the model. So, picking out the subset of significant features can be somewhat challenging. This task of identifying the best subset of features to include in the model, among all possible subsets of features, is referred to as *variable selection*.

10. Difference between R -squared versus adjusted R -squared

Both R^2 and Adjusted R^2 are used to evaluate the Performance or Quality of the model.

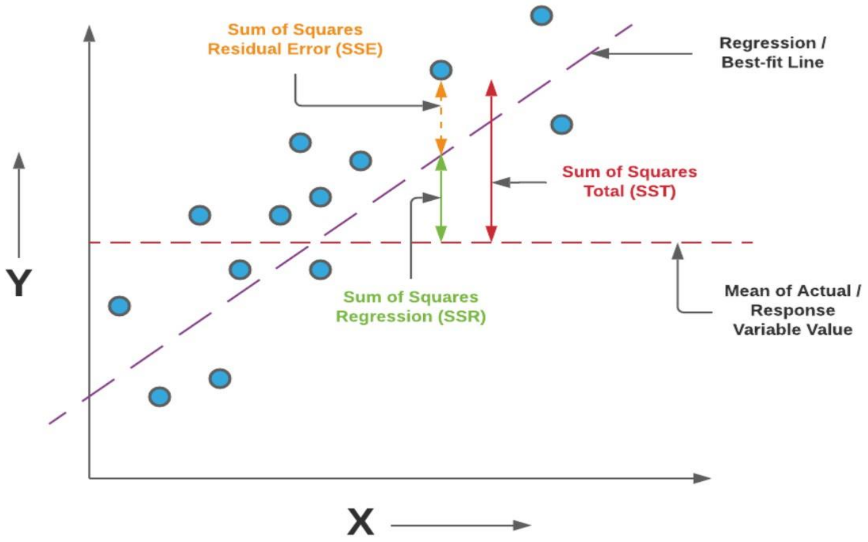
R^2

It enables us to compare our Model **Y_predicted** with constant baseline **Y_DAS**. i.e average or mean to determine the performance of the model

SS_{RES} = Sum of Square error or residuals

SS_{TOT} = Total sum of Square

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \overline{y})^2}$$



Adjusted R^2

Adjusted R-squared is a modified version of R-squared that has been adjusted for the number of predictors in the model. The adjusted R-squared increases when the new term improves the model more than would be expected by chance. It decreases when a predictor improves the model by less than expected. Typically, the adjusted R-squared is positive, not negative. It is always lower than the R-squared.

$$Adjusted\ R^2 = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1}$$

Where

R^2 Sample R-Squared

N Total Sample Size

p Number of independent variable