1 -> A

2 -> A

3 -> B

4 -> B

5 -> C

6 -> B

7 -> D

8 -> D

9 -> A

10 -> A

11 -> D

12 -> A

**13. Explain the term Regularization.**

Regularization is one of the most important concepts of machine learning. It is a technique to prevent the model from overfitting by adding extra information to it.

Sometimes the machine learning model performs well with the training data but does not perform well with the test data. It means the model is not able to predict the output when deals with unseen data by introducing noise in the output, and hence the model is called overfitted. This problem can be deal with the help of a regularization technique.

This technique can be used in such a way that it will allow to maintain all variables or features in the model by reducing the magnitude of the variables. Hence, it maintains accuracy as well as a generalization of the model.

It mainly regularizes or reduces the coefficient of features toward zero. In simple words, "*In regularization technique, we reduce the magnitude of the features by keeping the same number of features.*"

$$y= \beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\cdots+\beta_nx_n +b$$

In the above equation, Y represents the value to be predicted

X1, X2, …Xn are the features for Y.

β0,β1,…..βn are the weights or magnitude attached to the features, respectively. Here represents the bias of the model, and b represents the intercept.

Linear regression models try to optimize the β0 and b to minimize the cost function. The equation for the cost function for the linear model is given below:

$$\sum_{i=1}^{M}\left(y_i - y'_i\right)^2 = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{n} \beta_j * Xij\right)^2$$

Now, we will add a loss function and optimize parameter to make the model that can predict the accurate value of Y. The loss function for the linear regression is called as RSS or Residual sum of squares.

**14. Which particular algorithms are used for Regularization**

There are mainly two types of regularization techniques, which are given below:

- Ridge Regression
- Lasso Regression

Ridge Regression

- Ridge regression is one of the types of linear regression in which a small amount of bias is introduced so that we can get better long-term predictions.
- Ridge regression is a regularization technique, which is used to reduce the complexity of the model. It is also called as L2 regularization.
- In this technique, the cost function is altered by adding the penalty term to it. The amount of bias added to the model is called Ridge Regression penalty. We can calculate it by multiplying with the lambda to the squared weight of each individual feature.
- The equation for the cost function in ridge regression will be:

$$\sum_{i=1}^{M}\left(y_i - y'_i\right)^2 = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{n} \beta_j * x_{ij}\right)^2 + \lambda \sum_{j=0}^{n} {\beta_j}^2$$

- In the above equation, the penalty term regularizes the coefficients of the model, and hence ridge regression reduces the amplitudes of the coefficients that decreases the complexity of the model.
- As we can see from the above equation, if the values of λ tend to zero, the equation becomes the cost function of the linear regression model. Hence, for the minimum value of λ, the model will resemble the linear regression model.

- A general linear or polynomial regression will fail if there is high collinearity between the independent variables, so to solve such problems, Ridge regression can be used.
- It helps to solve the problems if we have more parameters than samples.

Lasso Regression:

- Lasso regression is another regularization technique to reduce the complexity of the model. It stands for Least Absolute and Selection Operator.
- It is similar to the Ridge Regression except that the penalty term contains only the absolute weights instead of a square of weights.
- Since it takes absolute values, hence, it can shrink the slope to 0, whereas Ridge Regression can only shrink it near to 0.
- It is also called as L1 regularization. The equation for the cost function of Lasso regression will be:

$$\sum_{i=1}^{M}(y_i - y'_i)^2 = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{n}\beta_j * x_{ij}\right)^2 + \lambda\sum_{j=0}^{n}|\beta_j|$$

- Some of the features in this technique are completely neglected for model evaluation.
- Hence, the Lasso regression can help us to reduce the overfitting in the model as well as the feature selection.

Key Difference between Ridge Regression and Lasso Regression

- Ridge regression is mostly used to reduce the overfitting in the model, and it includes all the features present in the model. It reduces the complexity of the model by shrinking the coefficients.
- Lasso regression helps to reduce the overfitting in the model as well as feature selection.

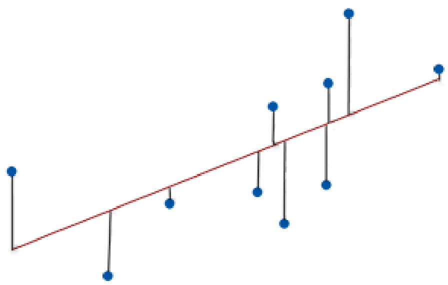
**15. explain the term error represent in linear regression equation.**

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

MSE is calculated by:

Mean squared error (MSE) measures the amount of error in statistical models. It assesses the average squared difference between the observed and predicted values. When a model has no error, the MSE equals zero. As model error increases, its value increases. The mean squared error is also known as the mean squared deviation (MSD).

For example, in regression, the mean squared error represents the average squared residual.



As the data points fall closer to the regression line, the model has less error, decreasing the MSE. A model with less error produces more precise predictions.

MSE Formula

The formula for MSE is the following.

$$MSE = \frac{\Sigma(y_i - \hat{y}_i)^2}{n}$$

Where:

- $y_i$ is the $i^{th}$ observed value.
- $\hat{y}_i$ is the corresponding predicted value.
- n = the number of observations.

The calculations for the mean squared error are similar to the variance. To find the MSE, take the observed value, subtract the predicted value, and square that difference. Repeat that for all observations. Then, sum all of those squared values and divide by the number of observations.

Notice that the numerator is the sum of the squared errors (SSE), which linear regression minimizes. MSE simply divides the SSE by the sample size.

Interpreting the Mean Squared Error

The MSE is the average squared distance between the observed and predicted values. Because it uses squared units rather than the natural data units, the interpretation is less intuitive.

Squaring the differences serves several purposes.

Squaring the differences eliminates negative values for the differences and ensures that the mean squared error is always greater than or equal to zero. It is almost always a positive value. Only a perfect model with no error produces an MSE of zero. And that doesn't occur in practice.

Additionally, squaring increases the impact of larger errors. These calculations disproportionately penalize larger errors more than smaller errors. This property is essential when you want your model to have smaller errors.

If you take the square root of the MSE, you obtain the root mean square error (RMSE), which does use the natural data units. In other words, MSE is analogous to the variance, whereas RMSE is akin to the standard deviation