Q1. R-squared or residual of square which one of these two is better measure of goodness of fit model in regression and why?

Ans. R-squared shows the fitness of data in regression model or it can be defined as "it shows the variance of dependent variables that can be explained by independent variables. The value of r-squared ranges between 0 and 1. If the value of r-squared is 0.9 that's mean the variance of dependent variables is explained by independent variables by 90%.

Residual some of square (RSS). RSS is a statistical technique which is used to measure the variance which is not explained by regression model. RSS=0 mean a perfect fit of the model.

If we compare these two techniques the r squared is considered the best fit for the goodness of the model because it provides the overall measure of the proportion of the variance in the dependent variable. The value of r-squared that ranges between 0 and 1 makes it easy to compare the fit of different models while the value of RSS cannot be easily compared across models because it depends on the scale of the dependent values.

Q2. What are TSS (total sum of square), ESS (explained sum of square) and RSS (residual sum of square)?

Ans. Total sum of square (TSS). In real world life the total sum of square is used to make the decisions related to investment. But to make these decisions requires a long history of past of that stock which mean collecting or adding more data points. In that case the sum of square will become larger, and the values will be more spreader out. In statistics it is used to calculate the variance and standard deviation. The sum of square is calculated as. TSS= $\sum_{i=0}^{n}(y_i - \bar{y})^2$.

Where $\bar{y}$ mean of all items and $(y_i - \bar{y})^2$ mean standard deviation of each item.

Residual sum of square (RSS). RSS is one of the kinds of TSS (total sum of square). The formula for RSS is. RSS= $\sum_{i=0}^{n}(y_i - \hat{y}_i)^2$ where

$\hat{y}_i$ value estimated by regression line

y_i is the observed value
sum of square explained the relation between variables, but the residual sum of square shows the unexplained the variability of variables. Its also shows the amount of error left in regression line. That is why it is also called the sum of square estimate of errors or SSE.

Explained sum of square (ESS). SSE is also a measure of how well the regression model fits the data. The fitness of data depends on the value of ESS. The higher value of ESS represents the fitness of model is not good. The formula for ESS is $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ where

$$\hat{y}_i = estimated\ value$$

$$\bar{y} = mean\ of\ the\ y$$

It is calculated by taking the square of difference of estimated value and mean of the value of each point and then by adding them up. The equation, which is used between TSS, RSS and ESS is TSS=RSS+ESS.

Q3. what is the need of regularization in machine learning?

Ans. In machine learning the models are often at the risk of overfitting and underfitting. To overcome this problem, we use regularization. This regularization reduces the model complexity and improve its performance on data set. L1, L2 and elastic net are the methods used in regularization. L1 automatically identify and discard irrelevant features. L2 adds square values of weights. Regularization is also important for neural network with complex and large models. It is also used in tree models and image classification.

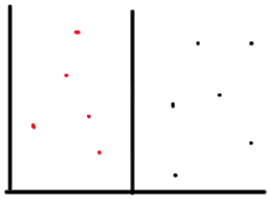
Regularization also makes the model less sensitive to small fluctuations. As a result, the data is more reliable and is more stable.

Q4. What is Gini-impurity index?

Ans. Gini-impurity index is a number between 0 and 0.5. It is used to measure the disorder in set of elements. The formula to calculate the Gini impurity index is

Gini impurity index=1-$\sum(P_i)^2$.

Where P_i is the probability of each class. If G=0 indicates the perfect pure node and 0.5 shows the impurity at the highest point. In Gini purity index we use the decision tree to classify the data. It starts from zero with same data label to impure the data to 1. In decision tree there will be two branches where it will purify the data into two parts by drawing a line in between.



Q5. Are unregularized decision trees prone to over fitting? If yes, why?

Ans. when machine learning tries to learn from details along with noise and fits the data to a curve is called over fitting. In more complex cases the over fitting occurs.

If we talk about the decision trees, then they handle both categorical and numerical data. Sometimes the decision trees are more sensitive to noise which cause the over fitting.

Poor generalization is another cause to overfitting in decision trees and they remain fail to make accurate decisions. Most of the time in complex and large cases the trees make the numerous branches which makes them harder to get meaningful information.

Q6. What is an ensemble technique in machine learning?

Ans. A combination of multiple models for prediction purpose is called ensemble technique. These techniques are used to improve the performance. A simple ensemble technique collect the prediction from different models and makes the final result ready in form of another prediction. This is a quiet easy way to compare the models.

Staking is an advanced ensemble technique. It makes a one meta model to make prediction which is the combination of multiple base models.

Averaging is also an ensemble method in machine learning. It takes the average of prediction from multiple models. This is commonly used in regression models. These advanced ensemble techniques are very effective because these techniques reduce the risks of over fitting and makes the generalization better.

Q7. Difference between bagging and boosting techniques?

Ans. Bagging and boosting both are two different ensemble methods.

Bagging algorithms can be parallelized for a better implementation. While the boosting is used to solve the complexity in data. Boosting handles both over fitting and under fitting.

In boosting the model work sequentially and improve the model predictions. The bagging is achieved through random sampling. This random sampling could be repeated.

In bagging multiple subset are prepared and base model is prepared on every subset. Boosting algorithm always try to correct the previous model error and the final model shows the weighted mean of all methods.

Q8. What is out of bag error in random forests?

Ans. out of bag or OOB is used for the accuracy in random forests. The random forests are popular because of accuracy. It has become a most common classification tool. The trees in random forests are constructed on random samples. It is usually a bootstrap sample and the data which is not the part of bootstrap sample is referred as the out of bag. So, OOB is used to estimate the error because it is calculated using the samples that are not used in the training of the models. So, it is the unbiased estimation of prediction of error.

Q9. What is k-fold cross validation?

Ans. Basically k-fold cross validation is the estimation of skill of model. In this method we have to choose the value of k, which must be chosen carefully for data samples. There is not formula rule to choose the value. Usually, it is 5 or 10. If the value of k is not chosen evenly in that case the data will split on one side which have only remainder. The k-fold cross validation ensure that it includes the data both from training set and testing set. It first identifies the groups in data set then split the groups in parts. The model is trained on k-1 part and then tested on remaining parts. This process is repeated k times. this methos works well with limited data.

Q10. What is hyperparameter tuning in ML and why it is done?

Ans. hyperparameter tuning helps to find the optimal sets of hyperparameter for a model. This process involves running multiple experiments with different sets of hyperparameters. While the hyperparameter are the values which can be tuned manually and automatically. There are several methods used in hyperparameter. Such as grid search, random search, and Bayesian optimization. grid search method of hyperparameter tuning includes the training model for every possible combination of hyperparameter. This is limited with the predefined set of possible values which may

not include the optimal values. Due to this simplicity, it is used for smaller models. Random search involves the randomly selecting combination of hyperparameter from predefined sets. It is also not effective to find the optimal values so cannot be used for large models. Bayesian optimization is used to find the optimal combination of hyperparameters in ML model. It can be used for large and complex models.

Q11. What issue can occur if we have large learning rate in Gradient descent?

Ans. As high learning rate helps the model to learn fast but it creates issues in large learning rate in gradient descent.

Overshooting happens where algorithm fails to converge to an optimal solution. It will skip the optimal solution.

A higher gradient value means lower learning rate while lower gradient mean the learning rate will be higher.

When learning rate is higher than the gradient descent may suffer from divergence. In this case the weights increase exponentially which results in instability and high loss of values.

Workload become unstable if the learning rate is too high.

High rate of learning increases the risks of overfitting which results in poor generalization performance on new data.

Q12. can we use logistic regression for classification of nonlinear data? If not, why?

Ans. As the logistic regression only takes linear relationship between input and output. That is why it is not useful for non-linear data.

Logistic regression is also sensitive to noise and outliers. Due to this drawback, it is not useful for non-linear models. It effects the accuracy and stability of the model

To learn the multiple features of models it does not has the capacity. So, it can only be used for linear regression models.

A13. Differentiate between adaboosting and gradient boosting?

Ans. Gradient boosting is more flexible while adaboosting only has few tuning parameters.

Gradient boosting can add new models to minimize the loss of function. Gradient boosting can handle complex models. Fradient boosting consider the sequentially when building the models. It achieves higher accuracy rate.

Here are few features of adaboosting which differentiate it from gradient boosting.

Adaboosting is much faster than gradient bosting when time is considered.

For simple models the adaboosting works better compared to gradient boosting.

The over fitting is better controlled in adaboosting.

Q14. What is bias variance trade off in ML?

Ans. the bias variance shows the relationship between the complexity of the model and prediction. As we increase the complexity of the model the variance decreases and its bias increases. It is used in financial forecasting, medical diagnostic and customer behaviour analysis.

Q15. Give short description each of linear, RBF and polynomial kernels used in SVM?

Ans. linear kernel. In linear kernel the data is divided inti two classes but using a single line the linear kernel methos algorithm find the best line to separate the data into two parts. It is the best way to classify the data into two equal classes.

RBF or radical base function. It is most common used function because of its flexibility. The RBF kernel calculate the similarity score between data points and assign them a close value point with each other. It is used for nonlinear dataset. The RBF kernel has two hyperparameter, c for SVM and y for RBF. RBF works by mopping the input data into higher dimensional feature space. We used the RBF when data is unevenly distributed.

Polynomial kernel. It creates a polynomial combination and generate the new features. The polynomial kernel has number of parameters that increase it performance. It maps the data into higher dimension space. It can also get the non-linear relation between input and output data. It can be combined with another kernel to handle even complex dataset. It can be used for image classification and text data classification it is used in bioinformatics and regression models.