Machine Learning

1.R-Squared or Residual sum of squares (RSS) which one of these two is better measure of goodness of fit model in regression and why?

Ans. R-squared is a better measure of goodness of fit model in regression than the RSS.

R-Squared is a statistical measure that represents the proportion of the variance for the dependent variable that is explained by the independent variable in the model.

It ranges from 0 to 1, where a value closer to 1 indicates a better fit.

Formula to calculate R-squared is :

$$R^2 = 1 - \frac{RSS}{TSS}$$

Where RSS is the residual sum of squares

And TSS is the total sum of squares

The reason why r-squared is often preferred over RSS is due to its standardized scales:

1. Scalability 2. Interpretability 3. Benchmarking 4. Adjustment for model complexity.

2. What are TSS, ESS and RSS in regression. Also mention the equation relating these 3 metrics with each other.

Ans. **TSS** (Total sum of squares) is the sum of square of difference of each data point from the mean value of all the values of target variable(y).

ESS (Explained sum of squares) is the sum of the differences between the predicted value and the mean of the dependent variable.

RSS (Residual sum of squares) is used to measure the amount of variance in a data set that is not explained by a regression model itself.

The relationship between TSS, RSS and ESS is as follows :

$$TSS = ESS + RSS$$

3. What is need of regularization in machine learning?

Ans. Regularization is crucial for addressing overfitting- where a model memorizes training data details but can't generalize to new data, and underfitting – where the model is too sample to capture the training data's complexity.

Regularization is used to prevent overfitting and improve the generalization performance of a model.

4. What is GIni – impurity index?

Ans. Gini index is a powerful measure of the randomness or the impurity or the entropy in the value of a dataset. Gini index aims to decrease the impurities from the root nodes to the leaf nodes of the decision tree model.

5. Are unregularized decision trees prone to overfitting? If yes, why?

Ans. Decision trees are popular and powerful method for data mining, as they can handle both numerical and categorical data and can easily interpret the results. However, decision trees can also suffer from overfitting, which means that they learn too much from the training data cand fail to generalize well to new data

6. What is ensemble technique in machine learning?

Ans. Ensemble technique aim at improving the accuracy of results in models by combining multiple models instead of using a single model. The combined models increase the accuracy of the results significantly.

7. What is the difference between bagging and boosting techniques?

Ans. Bagging and boosting are different ensemble techniques that use multiple models to reduce error and optimize the model

Bagging	**Boosting**
Bagging technique combines multiple models trained on different subsets of data	Boosting technique trains the model sequentially, focusing on the error made by the previous model

8. What is out-of-bag error in random forests?

Ans. Out-of-bag error (OOB) in random forest is the error rate of the model calculated on the samples not used during the training of each individual decision Tree.

It serves as internal estimate of the model's performance on unseen data without requiring a separate validation data.

9. What is k-fold cross validation?

Ans. In k-fold cross validation the data set is divided into a number of k-folds and used to assess the model's ability as new data becomes available.

K represents the number of groups into which the data sample is divided.

10. What is hyper parameter tuning in machine learning and why it is done?

Ans. Hyper parameters directly control model structure, function and performance. Hyper parameter tuning allows data scientists to tweak model performance for optimal results. The goal of hyper parameter tuning is to find the values that leads to the best performance on the given task. This process is an essential part of machine learning and choosing appropriate hyper parameter values is crucial for success.

11. What issues can occur if we have a large learning rate in gradient Descent?

Ans. If the learning rate is too high, the algorithm may overshoot the minimum and if the learning rate is too low, the algorithm may take too long to converge.

Gradient descent can overfit the training data if the model is too complex or the learning rate is too high.

12. Can we use logistic regression for classification of non-linear data? If not why?

Ans. Logistic regression is simple and easy to implement but it also has some drawback.

It assumes a linear relationship between the input features and the output. This means that it cannot capture the complexity and non-linearity of the data.

13. Differentiate between Adaboost and Gradient boosting.

Ans.

Adaboost	Gradient boosting
During each iteration in Adaboost, the weights of incorrectly classified samples are increased so that the next weak learner focuses more on these samples.	Gradient boosting updates the weights by computing the negative gradient of the loss function with respect to the predicted output.
Adaboost uses simple decision trees with one split known as the decision stumps of weak learners.	Gradient boosting can use a wide range of base learners such as decision trees, and linear models.
Adaboost is more susceptible to noise and outliers in the data as it assigns high weights to misclassified samples	Is generally more robust as it updates the weights based on the gradients which are less sensitive to outliers.
Algorithm: Training process starts with a decision tree stump(usually), at every step, the	Algorithm: GBM uses gradient descent to iteratively fit new weak learners to the

weight of the training samples which are misclassified are increased for next iteration. The next tree is built sequentially on the same training data but using the newly weighted training samples. This process is repeated until a desired performance is achieved	residuals of the previous ones, minimizing a loss function. There are several loss functions to choose from, Mean-squared-error being most common for Regression and Cross Entropy fro Classification. GMB uses decision tree as the weak learners
The final model is formed by combining the predictions from individuals trees through a weighted sum.	The final model is an equal-weighted sum of all the individual trees

14. What is bias-variance trade off in machine learning?

Ans. The bias variance trade off describes the relationship between a model's complexity, the accuracy of its predictions and how well it can make predictions on previously unseen data that were not used to train the model. In general as we increase the number of tunable parameters in a model, it becomes more flexible and can better fit a training data set. It is said to have lower error, or bias. However, for more flexible models, there will tend to be greater variance to the model fit each time we take a set of samples to create a new training data set. It is said that there is greater variance in the model's estimated parameters.

15. Give short description each of linear, RBF, polynomial kernels used in svm.

Ans. Linear kernel : It is used when the data is linearly separable.

$K(x1,x2)=x1.x2$

Radial-basis function kernel : It maps the input data to an infinite-dimensional space

$K(x,y) = exp(-y||x-y||^2)$

Polynomial kernels : It is used when the data is not linearly separable.

$K(x1,x2) = (x1.x2+1)d$