MACHINE LEARNING

1. **R-squared or Residual Sum of Squares (RSS) which one of these two is a better measure of goodness of fit model in regression and why?**

Ans - **R-squared is the one of these two is a better** measure of goodness of fit model in regression .

a higher r-squared indicates more variability is explained by the model.The best fit line is the one that minimises sum of squared differences between actual and estimated results. Taking average of minimum sum of squared difference is known as Mean Squared Error (MSE). Smaller the value, better the regression model

2. **What are TSS (Total Sum of Squares), ESS (Explained Sum of Squares) and RSS (Residual Sum of Squares) in regression. Also mention the equation relating these three metrics with each other.**

The total sum of squares (TSS or SST) tells you how far the data points in a dataset are from the center. It's a descriptive statistic called a measure of spread or dispersion. Dividing the TSS by the number of observations in the dataset gives you the average variability within the data, which is called the variance

Total Sum of Squares (TSS**)** expresses the total variation in Yi. (ignoring X) around the mean of Y: TSS = n.

The explained sum of squares (ESS) is the sum of the squares of the deviations of the predicted values from the mean value of a response variable, in a standard regression model

yi = a + b1x1i + b2x2i + ... + ei

The residual sum of squares (RSS) measures the level of variance in the error term, or residuals, of a regression model. The smaller the residual sum of squares, the better your model fits your data; the greater the residual sum of squares, the poorer your model fits your data.

SSR = S(– y)2 = SST – SSE.

the equation relating these three metrics with each other.

TSS = RSS + ESS

3. What is the need of regularization in machine learning?

the model can easily be overfitted or under fitted. To avoid this, we use regularization in machine learning to properly fit a model onto our test set. Regularization techniques help reduce the chance of overfitting and help us get an optimal 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."

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

Ridge Regression

Lasso Regression

4. What is Gini–impurity index?

Gini Impurity is a measurement used to build Decision Trees to determine how the features of a dataset should split nodes to form the tree. More precisely, the Gini Impurity of a dataset is a number between 0-0.5, which indicates the likelihood of new, random data being misclassified if it were given a random class label according to the class distribution in the dataset.

The Gini Index or Gini Impurity is calculated by subtracting the sum of the squared probabilities of each class from on

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

Yes, unregularized descision-trees prone to overfitting because they are very data intensive - that is, they examine the data in a lot of ways. At each node, they look at every possible split of every independent variable. sometimes they impose a rule of monotonicity if the variable is continuous or ordinal.

6. What is an ensemble technique in machine learning?

Ensemble methods is a machine learning technique that combines several base models in order to produce one optimal predictive model.

Ensemble methods are techniques that create multiple models and then combine them to produce improved results. Ensemble methods usually produces more accurate solutions than a single model would. This has been the case in a number of machine learning competitions, where the winning solutions used ensemble methods.

Voting and averaging are two of the easiest ensemble methods. They are both easy to understand and implement. Voting is used for classification and averaging is used for regression. In both methods, the first step is to create multiple classification/regression models using some training dataset.

7. What is the difference between Bagging and Boosting techniques?

Bagging and Boosting are two types of Ensemble Learning. These two decrease the variance of a single estimate as they combine several estimates from different models. So the result may be a model with higher stability.

Bagging: It is a homogeneous weak learners' model that learns from each other independently in parallel and combines them for determining the model average.

Boosting: It is also a homogeneous weak learners' model but works differently from Bagging. In this model, learners learn sequentially and adaptively to improve model predictions of a learning algorithm.

Bagging	Boosting
The simplest way of combining predictions belong to the different types.	A way of combining predictions that belong to the same type.
Aim to decrease variance, not bias.	Aim to decrease bias, not variance.
Bagging tries to solve the over-fitting problem.	Boosting tries to reduce bias

Each model receives equal weight.	Models are weighted according to their performance.

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

The out-of-bag (OOB) error is the average error for each calculated using predictions from the trees that do not contain in their respective bootstrap sample. This allows the RandomForestClassifier to be fit and validated whilst being trained

Bagging uses subsampling with replacement to create training samples for the model to learn from. OOB error is the mean prediction error on each training sample xi, using only the trees that did not have xi in their bootstrap sample.

Calculating out-of-bag error

The specific calculation of OOB error depends on the implementation of the model, but a general calculation is as follows.

1 Find all models (or trees, in the case of a random forest) that are not trained by the OOB instance.

2 Take the majority vote of these models' result for the OOB instance, compared to the true value of the OOB instance.

3 Compile the OOB error for all instances in the OOB dataset.

9. What is K-fold cross-validation?

Cross-validation is a statistical method used to estimate the skill of machine learning models.

It is commonly used in applied machine learning to compare and select a model for a given predictive modeling problem because it is easy to understand, easy to implement, and results in skill estimates that generally have a lower bias than other methods.

It helps us to avoid overfitting. As we know when a model is trained using all of the data in a single short and give the best performance accuracy. To resist this k-fold cross-validation helps us to build the model is a generalized one

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

Hyper parameter tuning consists of finding a set of optimal hyper parameter values for a learning algorithm while applying this optimized algorithm to any data set. That combination of hyper parameters maximizes the model's performance, minimizing a predefined loss function to produce better results with fewer errors.hyperparameter-tuning is important to find the possible best sets of hyper parameters to build the model from a specific dataset.

The values of hyperparameters might improve or worsen your model's accuracy

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

In order for Gradient Descent to work, we must set the learning rate to an appropriate value. This parameter determines how fast or slow we will move towards the optimal weights. If the learning rate is very large we will skip the optimal solution.

Gradient Descent is too sensitive to the learning rate. If it is too big, the algorithm may bypass the local minimum and overshoot and it can cause undesirable divergent behavior in your loss function

12. Can we use Logistic Regression for classification of Non-Linear Data? If not, why?

Logistic Regression has traditionally been used as a linear classifier, i.e. when the classes can be separated in the feature space by linear boundaries.

Logistic regression is known and used as a linear classifier. It is used to come up with a hyperplane in feature space to separate observations that belong to a class from all the other observations that do not belong to that class. The decision boundary is thus linear. Robust and efficient implementations are readily available (e.g. scikit-learn) to use logistic regression as a linear classifier.

13. Differentiate between Adaboost and Gradient Boosting.

S.No	Adaboost	Gradient Boost
1	An additive model where shortcomings of previous models are identified by high-weight data points	An additive model where shortcomings of previous models are identified by the gradient
2	The trees are usually grown as decision stumps	The trees are grown to a greater depth usually ranging from 8 to 32 terminal nodes.
3	Each classifier has different weights assigned to the final prediction based on its performance.	All classifiers are weighed equally and their predictive capacity is restricted with learning rate to increase accuracy.
4	It gives weights to both classifiers and observations thus capturing maximum variance within data.	It builds trees on previous classifier's residuals thus capturing variance in data.

14.what is bias variance tradeoff in machine learning?

In statistics and machine learning, the bias–variance tradeoff is the property of a model that the variance of the parameter estimated across samples can be reduced by increasing the bias in the estimated parameters.

It helps optimize the error in our model and keeps it as low as possible. An optimized model will be sensitive to the patterns in our data, but at the same time will be able to generalize to new data. In this, both the bias and variance should be low so as to prevent overfitting and underfitting

15. Give short description each of Linear, RBF, Polynomial kernels used in SVM.

SVM algorithms use a set of mathematical functions that are defined as the kernel. The function of kernel is to take data as input and transform it into the required form. Different SVM algorithms use different types of kernel functions. These functions can be different types

Kernel functions for sequence data, graphs, text, images, as well as vectors. The most used type of **kernel function is RBF.** Because it has localized and finite response along the entire x-axis.

RBF is the default kernel used within the sklearn's SVM classification algorithm

Linear

These are commonly recommended for text classification because most of these types of classification problems are linearly separable.

The linear kernel works really well when there are a lot of features, and text classification problems have a lot of features. Linear kernel functions are faster than most of the others and you have fewer parameters to optimize.

Polynomial

The polynomial kernel isn't used in practice very often because it isn't as computationally efficient as other kernels and its predictions aren't as accurate.