MACHINE LEARNING

Q1 to Q15 are subjective answer type questions, Answer them briefly.

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 (R^2) is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model. On the otherhand, residual sum of squares (RSS) is a statistical technique used to measure the amount of variance in a data set that is not explained by a regression model itself. Instead, it estimates the variance in the residuals, or error term. Thus, we can say that the RSS is the absolute amount of explained variation whereas R-squared is the absolute amount of variation as a proportion of total variation.

R-squared is a better measure for goodness of fit model in regression because even if there are large negative values (which are points far below the line) as similar to the large positive ones (points which are high above the line). So, by squaring the residual values we can treat both positive and negative discrepancies in the same manner.

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.

Ans.)

TSS (Total Sum of Squares)	ESS (Explained Sum of Squares)	RSS (Residual Sum of Squares)
The total sum of squares measures the deviation of data points away from the mean value. A higher sum of squares indicates higher variability while a lower result indicates low variability from the mean.	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.	Residual sum of squares (RSS) is a statistical technique used to measure the amount of variance in a data set that is not explained by a regression model itself. Instead, it estimates the variance in the residuals, or error term.
Total sum of squares (TSS) = explained sum of squares (ESS) + residual sum of squares (RSS).		

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

Ans.) Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting.It helps to improve the accuracy of the model.

4. What is Gini–impurity index?

Ans.) Gini Impurity Index is used to calculate the amount of probability of a specific feature that is classified incorrectly when selected randomly. It is referred as the measurement used to build Decision Trees to determine how the features of a dataset should split nodes to form the tree.

Gini impurity is calculated by subtracting the sum of the squared probabilities of each class from one.

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

Ans.) Yes. decision trees are prone to overfitting, especially when a tree is particularly deep. This is due to the amount of specificity we look at leading to smaller sample of events that meet the previous assumptions. This small sample could lead to unsound conclusions. This affects the accuracy of the predictions.

6. What is an ensemble technique in machine learning?

Ans.) 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.

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

Ans.) Bagging is the simplest way of combining predictions that belong to the same type while Boosting is a way of combining predictions that belong to the different types. Bagging aims to decrease variance, not bias while Boosting aims to decrease bias, not variance.

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

Ans.) 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.

9. What is K-fold cross-validation?

Ans.) K-fold Cross-Validation is when the dataset is split into a K number of folds and is used to evaluate the model's ability when given new data. K refers to the number of groups the data sample is split into. For example, if you see that the k-value is 5, we can call this a 5-fold cross-validation.

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

Ans.) Hyperparameter tuning consists of finding a set of optimal hyperparameter values for a learning algorithm while applying this optimized algorithm to any data set. That combination of hyperparameters maximizes the model's performance, minimizing a predefined loss function to produce better results with fewer errors.

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

Ans.) Gradient Descent is too sensitive to the learning rate. If it is too big, the algorithm may bypass the local minimum and overshoot.

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

Ans.) No, we cannot use Logistic Regression for Non-Linear data because it has a linear decision surface.

13. Differentiate between Adaboost and Gradient Boosting.

Ans.)

Adaboost	Gradient Boosting
AdaBoost is the first designed boosting algorithm with a particular loss function.	Gradient Boosting is a generic algorithm that assists in searching the approximate solutions to the additive modelling problem.
It minimises the exponential loss function that can make the algorithm sensitive to the outliers.	Any differentiable loss function can be utilised. Infact it is more robust to outliers than AdaBoost.
The method was mainly designed for binary classification problems and can be utilised to boost the performance of decision trees.	The technique can be used for both classification and regression problems.

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

Ans.) In 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.

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

Ans.) **Linear Kernel** - It is one of the most common kernels to be used. It is mostly used when there are a Large number of Features in a particular Data Set. Linear Kernel is used when the data is Linearly separable, that is, it can be separated using a single Line.

RBF Kernel - Radial Basis Networks are simple two-layer architectures with one layer of RBF neurons and one layer of output neurons. RBF kernels place a radial basis function centered at each point, then perform linear manipulations to map points to higher-dimensional spaces that are easier to separate.

Polynomial Kernel - The polynomial kernel represents the similarity of vectors (training samples) in a feature space over polynomials of the original variables, allowing learning of non-linear models.