MACHINE LEARNING

1) R-squared is generally considered a more comprehensive measure of goodness of fit compared to RSS, mainly because,
   R-squared provides a clear interpretation of the proportion of variance explained.
   R-squared allows for easier comparison across models.
   R-squared is normalized, which helps in comparing the goodness of fit across different datasets and models.

2) Total Sum of Squares (TSS) is the sum of squared differences between the observed dependent variables and the overall mean. It is the dispersion of the observed variable around the mean.

   Explained Sum of Squares (ESS) is the sum of the differences between the predicted value and the mean of the dependent variable. It describes how well our line fits to data.

   The Residual Sum of Squares (RSS) is the difference between the observed and predicted values.

   Equation = SST = SSR + SSE.

3) Regularization helps control model complexity by preventing overfitting to training data, resulting in better generalization to new data.
   One way to prevent overfitting is to use regularization, thereby preventing a model from becoming overly complex and memorizing the training data instead of learning its underlying patterns.
   Regularized models learn underlying patterns of data for better generalization to new data, instead of memorizing specific examples.
   Regularization can help balance the trade-off between model bias (underfitting) and model variance (overfitting) in machine learning, which leads to improved performance.
   When features are highly correlated, regularization can stabilize the model by reducing coefficient sensitivity to small data changes.

4) The Gini Impurity Index is a method used to measure the impurity of a dataset.
   It indicates the number of distinct classes present within a subset, with higher values indicating greater impurity. In decision tree algorithms, the Gini Index helps determine the best feature to split the data on at each node of the tree.

5) Yes, unregularized decision trees are prone to overfitting because,
   - Decision trees can overfit when there is limited training data.
   - As decision trees grow deeper, they become more susceptible to overfitting. Which increase complexity of tree.
   - As the tree becomes deeper, it not only captures relevant patterns but also memorizes noise, outliers, and specific anomalies present in the training set, leading to overfitting.

6) Ensemble learning is a machine learning technique that combines the predictions from multiple individual models to obtain a better predictive performance than any single model.

   Ensemble learning technique in machine learning.
   - Gradient Boosting Machines (GBM)
   - Extreme Gradient Boosting (XGBoost)
   - Cat Boost.
   - Stacking.
   - Random Subspace Ensembles.
   - Random Forest Variants.

7)

Bagging	Boosting
It is a homogeneous weak learners' model that learns from each other independently in parallel and combines them to determine the model average.	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.
The original dataset is divided into multiple subsets, selecting observations with replacements.	The new subset contains the components mistrained by the previous model.
This method combines predictions that belong to the same type.	This method combines predictions that belong to different types.
Bagging decreases variance.	Boosting decreases bias.
Base classifiers are trained parallelly.	Base classifiers are trained sequentially.
The models are created independently.	The model creation is dependent on the previous ones.

8) The out-of-bag error is the average error for each predicted outcome calculated using predictions from the trees that do not contain that data point in their respective bootstrap sample.

9) K-fold cross-validation is a widely used technique in machine learning and statistics for evaluating the performance and generalization ability of a predictive model.
   It provides a robust estimate of a model's performance by using different subsets of the data for training and validation.

10) Hyperparameter tuning is the practice of identifying and selecting the optimal hyperparameters for use in training a machine learning model. It is done because
    hyperparameter tuning minimizes the loss function of a machine learning model, which means that the model performance is trained to be as accurate as possible.
    it lays the groundwork for a model's structure, training efficiency, and performance. Optimal hyperparameter configurations lead to strong model performance.

11) The following issues can occur if we have a large learning rate in Gradient Descent The algorithm may skip the optimal solution.
    - The algorithm may skip the optimal solution.
    - The algorithm may fail to converge.
    - Frequent updates can lead to erratic convergence.
    - Careful tuning of the learning rate is required.
    - Large learning rates can cause weights to diverge.

12) Logistic Regression is primarily designed for linear classification. However, it can be adapted for non-linear data classification through some techniques:
    - Kernal Trick
    - Non-Linear Model
    - Feature Engineering.

13)

Adaboost(Adaptive Boosting)	Gradient Boosting
Adaboost is a popular ensemble learning algorithm that aims to improve the performance of machine learning models by combining multiple weak classifiers into a single strong classifier	Gradient Boosting is an advanced ensemble machine learning technique that combines multiple weak learners to form a strong predictive model
AdaBoost or Adaptive Boosting is the first Boosting ensemble model.	Gradient Boost is a robust machine-learning algorithm made up of Gradient descent and Boosting.
In AdaBoost, the weights of the samples are adjusted at each iteration.	No reweighting of the samples takes place in Gradient Boosting.

14) The bias-variance trade-off is a fundamental concept in machine learning that describes the balance between two types of errors that can affect the performance of a model, bias and variance.

15) Linear Kernel- The linear kernel is the simplest kernel function used in SVM. It computes the dot product between two input vectors, effectively measuring their similarity in the feature space.

    - Best suited for linearly separable data.
    - Simple and efficient for high-dimensional datasets.

    RBF kernel- The RBF Kernel also known as the Gaussian kernel, is a non-linear kernel that maps input vectors into an infinite-dimensional space.
    - Suitable for non-linear data where relationships are complex and not linearly separable.
    - High flexibility with γ parameter tuning.

- Effective in capturing non-linear relationships

16) Polynomial Kernel –

The polynomial kernel is a non-linear kernel that represents the similarity of input vectors in a polynomial feature space. It is an extension of the linear kernel, allowing for more complex decision boundaries by introducing polynomial terms.

- Suitable for scenarios where interactions between features are expected to be polynomial in nature.
- Used in natural language processing tasks and pattern recognition.
- Flexibility to model complex patterns with an adjustable polynomial degree

MCQ

1) D) Expected.
2) C) Frequencies.
3) C) 6.
4) B) Chisqared distribution.
5) C) F Distribution.
6) B) Hypothesis.
7) A) Null Hypothesis.
8) A) Two tailed.
9) B) Research Hypothesis.
10) A) NP.