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?

R-squared is preferred as it gauges the proportion of variance explained by the model, crucial for assessing goodness of fit. It quantifies the model's effectiveness in capturing the variability of the dependent variable, offering a more comprehensive evaluation. RSS solely focuses on unexplained variability, overlooking the portion of variance accounted for by the model, rendering it less informative for assessing model fit in regression analysis.

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.

TSS represents the total variability in the dependent variable, while ESS quantifies the variability explained by the model. RSS reflects the unexplained variability remaining after model fitting. These metrics are interconnected through the equation: TSS = ESS + RSS. Understanding these measures aids in comprehending the distribution of variability in the data and the efficacy of the regression model in capturing this variability

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

Regularization serves as a critical tool in mitigating overfitting by imposing penalties on excessively complex models. It encourages simpler models by constraining the magnitude of coefficients, thereby preventing overemphasis on noisy or irrelevant features. This regularization technique enhances model generalization to unseen data, promoting robust performance in machine learning tasks across diverse datasets and scenarios.

4. What is Gini–impurity index?

The Gini impurity index measures the impurity or disorder within a set of elements in a decision tree. It quantifies the probability of incorrectly classifying a randomly chosen element based on the distribution of classes within the set. A lower Gini impurity indicates a more homogeneous set, while higher values suggest greater impurity and randomness.

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

Yes, unregularized decision trees are susceptible to overfitting due to their capability to grow very deep to fit the training data perfectly. This excessive growth allows the model to capture noise and outliers present in the training data, leading to poor generalization performance on unseen data. Regularization techniques such as pruning or limiting the tree depth can mitigate this issue by promoting simpler tree structures.

6. What is an ensemble technique in machine learning?

Ensemble techniques in machine learning combine multiple individual models to improve prediction accuracy and robustness. By leveraging the wisdom of crowds, ensemble methods mitigate the weaknesses of individual models, resulting in better overall performance. Common ensemble techniques include bagging, boosting, and stacking.

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

Bagging and boosting are both ensemble techniques used to improve model performance by combining multiple weak learners. However, they differ in their approach. Bagging builds multiple models independently from random subsets of the training data and combines their predictions through averaging or voting. Boosting, on the other hand, builds models

sequentially, with each subsequent model focusing on the errors made by the previous ones, thereby reducing the overall error.

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

The out-of-bag (OOB) error in random forests is an estimate of the model's performance on unseen data. In random forest training, each decision tree is trained on a bootstrap sample of the data, leaving out around one-third of the data points on average. The OOB error is calculated by evaluating each data point using only the trees that were not trained on it, providing an unbiased estimate of the model's generalization error without the need for a separate validation set.

9. What is K-fold cross-validation?

K-fold cross-validation is a technique used to assess the performance of a machine learning model. It involves splitting the dataset into K subsets of equal size, where each subset is used as a validation set exactly once, with the remaining K-1 subsets used for training. This process is repeated K times, and the performance metrics are averaged across all folds to obtain an overall estimate of the model's performance. K-fold cross-validation provides a robust assessment of a model's generalization ability and helps prevent overfitting by utilizing the entire dataset for both training and validation.

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

Hyperparameter tuning in machine learning involves the process of selecting the optimal values for the hyperparameters of a model to maximize its performance on unseen data. Hyperparameters are parameters that are not learned during the training process but are set prior to training. Examples include the learning rate in gradient descent, the depth of a decision tree, or the number of hidden layers in a neural network. Hyperparameter tuning is essential

because the performance of a model can vary significantly with different hyperparameter values, and selecting the right values can greatly improve the model's predictive accuracy.

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

Large learning rates in gradient descent can lead to convergence issues such as overshooting the minimum or oscillations around the minimum. Overshooting occurs when the steps taken during optimization are too large, causing the algorithm to miss the minimum and diverge. Oscillations can occur when the learning rate is too high, causing the algorithm to bounce back and forth across the minimum, making convergence slow or impossible. In both cases, the optimization process becomes unstable, and the algorithm fails to find the optimal solution.

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

Logistic Regression is inherently a linear classifier, meaning it assumes a linear relationship between the independent variables and the log-odds of the dependent variable. Therefore, it cannot effectively capture the non-linear relationships present in the data. While it can still be applied to non-linear data, it may not perform as well as other algorithms specifically designed for non-linear classification tasks, such as decision trees or support vector machines with non-linear kernels.

13. Differentiate between Adaboost and Gradient Boosting.

Adaboost	**Gradient Boosting**
Sequentially trains a series of weak learners, with each subsequent learner focusing on the mistakes made by the previous ones.	Fits each weak learner to the residual errors of the previous learner, effectively minimizing the errors at each step.
Assigns weights to training instances, with higher weights placed on incorrectly classified instances to improve subsequent model iterations.	Does not assign weights to training instances but focuses on minimizing the overall error by fitting subsequent models to the residuals of the previous models.
Often uses decision trees with limited depth as weak learners, such as stump classifiers (decision trees with a single node and two leaves).	Can use a variety of weak learners, including decision trees, regression models, or other base learners.
Tends to be more susceptible to noisy data and outliers due to its iterative nature, as it may excessively focus on difficult-to-classify instances.	Generally,more robust to noisy data and outliers, as subsequent models are trained to correct the errors of previous models, leading to improved overall performance.
May converge faster than gradient boosting but can be more prone to overfitting if the number of weak learners is too high or if the weak learners are too complex.	Often slower to converge than Adaboost but tends to have better generalization performance, especially when using regularization techniques or simpler weak learners.

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

The bias-variance trade-off is a fundamental concept in machine learning that refers to the balance between bias and variance in the performance of a model. Bias measures the error introduced by approximating a real-world problem with a simplified model, while variance measures the model's sensitivity to fluctuations in the training data. A high-bias model tends to underfit the data, while a high-variance model tends to overfit the data. Finding the right balance between bias and variance is essential for building models that generalize well to unseen data.

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

1. Linear kernel: Computes the dot product of input samples in the original feature space, suitable for linearly separable data.

2. RBF (Radial Basis Function) kernel: Maps samples into a higher-dimensional space based on their similarity to landmarks, allowing for non-linear decision boundaries.

3. Polynomial kernel: Computes the dot product of input samples in a polynomial feature space, enabling the model to capture non-linear relationships between features.

STATISTICS WORKSHEET-5

Q1 to Q10 are MCQs with only one correct answer. Choose the correct option.

1. Using a goodness of fit,we can assess whether a set of obtained frequencies differ from a set of frequencies.
   a) Mean
   b) Actual
   c) Predicted
   d) **Expected**
2. Chisquare is used to analyse
   a) Score
   b) Rank
   c) Frequencies
   d) **All of these**
3. What is the mean of a Chi Square distribution with 6 degrees of freedom?
   a) 4
   b) 12
   c) **6**
   d) 8
4. Which of these distributions is used for a goodness of fit testing?
   a) **Normal distribution**
   b) Chisqared distribution
   c) Gamma distribution
   d) Poission distribution
5. Which of the following distributions is Continuous
   a) Binomial Distribution
   b) Hypergeometric Distribution
   c) **F Distribution**
   d) Poisson Distribution
6. A statement made about a population for testing purpose is called?
   a) Statistic
   b) **Hypothesis**
   c) Level of Significance
   d) TestStatistic
7. If the assumed hypothesis is tested for rejection considering it to be true is called?
   a) **Null Hypothesis**
   b) Statistical Hypothesis
   c) Simple Hypothesis
   d) Composite Hypothesis
8. If the Critical region is evenly distributed then the test is referred as?
   a) **Two tailed**
   b) One tailed
   c) Three tailed
   d) Zero tailed
9. Alternative Hypothesis is also called as?
   a) Composite hypothesis
   b) **Research Hypothesis**
   c) Simple Hypothesis
   d) Null Hypothesis

Statistic Worksheet 5

10. In a Binomial Distribution, if ‘n’ is the number of trials and ‘p’ is the probability of success, then the mean value is given by _____
    a) **np**
    b) n