**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?**

R-squared is generally considered a better measure of goodness of fit in regression models compared to the Residual Sum of Squares (RSS). It measures the proportion of the variance in the department variable that is explain by the independent variables in the models. It ranges from 0-1 with 1 indicating a perfect fit. Higher values of R-squared suggestion that the model explains a larger portion of the variance in the dependent variable.

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

In regression analysis, TSS , ESS and RSS are metrices used yo assess the variance and goodness of the fit of a regression model.

1.Total sum of squares (TSS)- It measures the total variability of in the dependent variable (Y) before accounting for the regression modes. It quantifies the total deviation of each point from the mean of the dependent variable.

TSS=∑(*Yi*​−*Y*ˉ)2

2.Explained sum of squares (ESS)-It quantifies the variability in the dependent variable that is explained by the regression model. It measures the deviation of the predicted values from the mean of the dependent variable.

ESS=∑(*Y*^*i*​−*Y*ˉ)2 where *Yi*​ is the predicted value of the dependent variable from the regression model, and *Y*ˉ is the mean of the dependent variable.

3.Residual Sum of squares-It measures the unexplained variability in the dependent variable by the regression model. It quantifies the deviation of the observed values from the predicted values.

RSS- ∑(*Yi*​−*Y*^*i*​)2 where *Yi*​ is the observed value of the dependent variable, and *Y*^*i*​ is the predicted value of the dependent variable from the regression model.

The relationship between these metrics can be expressed as: TSS=ESS+RSSTSS=ESS+RSS This equation demonstrates that the total variability (TSS) in the dependent variable can be partitioned into the variability explained by the regression model (ESS) and the unexplained variability (RSS).

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

Regularization in machine learning is used to prevent overfitting and improve the generalization of models. Overfitting occurs when a model learns not only the underlying patterns in the training data but also learns noise or random fluctuations, which can negatively impact its performance on unseen or new data.

**4. What is Gini–impurity index?**

The Gini impurity index is a measure used in decision tree algorithms, particularly in the context of binary classification problems. It quantifies the impurity or disorder within a set of class labels in a dataset.

The Gini impurity of a set measures the probability of incorrectly classifying a randomly chosen element in that set if it were randomly labeled according to the distribution of labels in the subset.

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

Yes, unregularized decision trees are prone to overfitting, and there are several reasons for this:

1.High variance

2.Capturing noise

3.overly complex tree

4.Sensitive to smell variations

**6. What is an ensemble technique in machine learning?**

Ensemble techniques in machine learning involve combining multiple individual models to create a more powerful, robust, and accurate predictive model. Instead of relying on a single model's prediction, ensemble methods leverage the diversity and collective wisdom of multiple models to make better predictions.

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

**Bagging**: It involves training multiple instances of the same model on different subsets of the training data, often using bootstrap sampling. For example, Random Forest is an ensemble of decision trees built through bagging, where each tree is trained on a random subset of the data and features.

**Boosting**: Boosting techniques sequentially train models and assign higher weights to misclassified instances to improve prediction accuracy. Popular algorithms like AdaBoost, Gradient Boosting Machines (GBM), and XGBoost use boosting principles.

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

The out-of-bag (OOB) error is an evaluation method used specifically in Random Forests, a type of ensemble learning algorithm based on decision trees.

In Random Forests, each tree in the forest is built using a bootstrap sample (sampling with replacement) from the original dataset. During this process, some data points are not included in the bootstrap sample for each tree. These data points that are not sampled for a particular tree are referred to as the out-of-bag (OOB) samples for that tree.

**9. What is K-fold cross-validation?**

**K-fold cross-validation is a resampling technique used to evaluate machine learning models on a limited dataset.** It aims to provide a more accurate and reliable estimate of a model's performance on unseen data compared to a single train-test split.

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

Hyperparameter tuning refers to the process of selecting the best set of hyperparameters for a machine learning model. Hyperparameters are configuration settings for a model that are not learned from the data during training but are set before the training process begins. Examples include learning rate, regularization strength, tree depth, and kernel type.

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

A large learning rate in Gradient Descent can lead to several issues:

Divergence: A large learning rate can cause the optimization process to overshoot the minimum point or bounce around it, preventing convergence. The algorithm may fail to find the minimum and diverge, resulting in unstable or exploding gradients.

Overshooting the minimum: Large steps taken by the algorithm can cause it to overshoot the optimal point. Instead of converging towards the minimum, the algorithm may oscillate or move too far, making it difficult to reach convergence.

**Waste of resources:** Training with a large learning rate takes longer to converge compared to a smaller one. This can be computationally expensive and waste time especially when dealing with large datasets

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

Logistic Regression is a linear classification algorithm, meaning it creates a linear decision boundary between classes. This linear decision boundary makes Logistic Regression suitable for linearly separable data where classes can be divided by a straight line, plane, or hyperplane.

**13. Differentiate between Adaboost and Gradient Boosting.**

Adaboost-Focuses on sequential learning, where subsequent models correct the errors made by previous models. Adjusts weights on misclassified data points to emphasize them in subsequent iterations.Learns from errors by increasing the importance of misclassified samples, allowing subsequent models to focus more on these instances.

Gradient Boosting - Works by sequentially fitting models to the residuals (errors) of the previous models. Each new model focuses on minimizing the errors made by the previous

ensembles.Gradient Boosting focuses on residuals, fitting new models to the errors of previous ones, using a gradient descent approach.

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

The bias-variance tradeoff is a fundamental concept in machine learning that describes the relationship between a model's complexity, its ability to capture underlying patterns in data, and its generalization performance on unseen data.

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

1.Linear Kernel-Simplest kernel: It computes a direct dot product between data points.Suitable for linearly separable data: When a clear decision boundary can be drawn as a straight line.

Advantages: Computationally efficient, easy to interpret.

Disadvantages: Limited ability to capture complex patterns, may not be effective for nonlinearly separable data.

2. Radial Basis Function (RBF) Kernel:Nonlinear kernel: Projects data into a higher-dimensional space to find nonlinear decision boundaries.Measures similarity based on distance between points: Assigns higher similarity to closer points.Versatile: Handles various datasets and often performs well without extensive hyperparameter tuning.

Advantages: Effective in handling nonlinear relationships, less prone to overfitting compared to polynomial kernels.

Disadvantages: Can be computationally expensive for large datasets, requires careful tuning of the gamma parameter (which controls the kernel's width).

3. Polynomial Kernel:Nonlinear kernel: Creates more complex decision boundaries by considering combinations of features.Calculates similarity using a polynomial function of the dot product: Allows for modeling of interactions between features.

Advantages: Can model complex relationships, can be useful for problems with polynomial feature interactions.

Disadvantages: More prone to overfitting, requires careful tuning of the degree parameter (which controls the complexity of the polynomial).