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 the goodness of fit model in regression and why?

R-squared is commonly used because it provides an overall assessment of how well the model explains the variability of the dependent variable. However, RSS provides additional insights, especially model comparison, and assessment of model complexity. It is advised to use both measures to get a comprehensive understanding of the model's performance.

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.

Total Sum of Squares (TSS):

Definition: Quantifies how much the observed data points deviate from the mean.

Equation:

TSS = $\sum_{i=1}^{n}(y_i - \bar{y})^2$

n = number of observations

y_i= value for each data point

$\bar{y}$ = mean of the data set

Explained Sum of Squares (ESS):

Definition: Represents the deviation between the predicted values (from the regression line) and the mean of the dependent variable.

Equation:

ESS = $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$

n = number of observations

$\hat{y}_i$ = predicted values for individual data point

$\bar{y}$ = mean of the dependent variable

Residual Sum of Squares (RSS):

Definition: Represents the difference between the observed values and the predicted values (residuals).

Equation:

RSS = $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

n = number of observations

$\hat{y}_i$ = predicted values of each data point

y_i = observed value of dependent variable

Relationship Between TSS, ESS, and RSS:

TSS = ESS + RSS

TSS represents the total variation in the data. ESS represents the variation explained by the regression model. RSS represents the unexplained variation (residuals).

R-squared (Coefficient of Determination):

R-squared is derived from the three measures. It quantifies the proportion of variability in the dependent variable that the model explains and a higher R^2 indicates a better fit of the model to the data.

Equation:

R^2 = 1- $\frac{RSS}{TSS}$ therefore $R^2 = \frac{ESS}{TSS}$

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

Regularization in machine learning prevents overfitting or underfitting by enhancing model performance and generalization. Hence reducing errors.

4. What is Gini–impurity index?

Gini-impunity is a metric used to build decision trees to determine how the features of a dataset should split nodes to form the tree. The index ranges from 0 to 0.5, lower values indicate less impurity and higher values more impurity.

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

Yes, unregularized decision- trees are prone to overfitting because they do not generalize well to new data.

6. What is an ensemble technique in machine learning?

An ensemble technique in machine learning is a method that aims at improving the accuracy of results in models by combining multiple models instead of a single model.

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

Bagging aims to reduce variance and overfitting by averaging predictions from multiple independently trained models while boosting reduces bias by iteratively improving model performance based on the errors made by previous models.

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

Out-of-bag error evaluates the model's performance on unseen data points that were not included in the training set for individual trees, it provides an unbiased estimate of how well the model will perform on new data.

9. What is K-fold cross-validation?

K-fold cross-validation is a technique for evaluating predictive models. The data set is divided into k subsets (or folds), and the model is trained and tested on different subsets of data, reducing the risk of bias. It maximizes the usage of available data for both training and testing purposes.

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

Hyperparameter tuning selects optimal pre-training parameter values, crucial for performance optimization by adjusting factors like learning rates or tree depths. It prevents overfitting and improves a model's generalization to new data.

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

Overshooting, instability, poor convergence, and difficulty in fine-tuning could occur.

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

Logistic regression is commonly used as a linear classifier. However, linear data doesn't adhere to a linear relationship between features and the target. The decision boundary needed to separate different classes may be curved or more complex than what a linear model like logistic regression can handle.

13. Differentiate between Adaboost and Gradient Boosting.

AdaBoost trains weak learners sequentially, giving more weight to misclassified instances to improve classification performance, especially in challenging instances. Gradient Boosting builds weak learners stage-wise, each correcting the errors of the previous model by fitting them to residual errors without adjusting instance weights. They are both ensemble techniques which differ in their approach to training, loss function optimization, handling of misclassified instances and robustness to outliers.

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

Bias-variance trade off in machine learning describes the balance between the simplicity of a model (bias) and its ability to capture complex patterns (variance). Models with high bias may underfit the data, while those with high variance may overfit. Achieving optimal performance involves finding the right balance between bias and variance, often through selecting appropriate model complexity and regularization techniques.

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

Linear Kernels:

It is the simplest kernel for SVM that computes the dot product of the input data points in the original feature space. It works well when the data is linearly separably or nearly separable.

RBF (Radical Basis Function) Kernels:

It is the most powerful and widely used kernel. It maps the data into high dimensional space using the radial basis function. It is effective for handling non-linearly separable data due to its flexibility.

Polynomial Kernels:

It introduces nonlinearity by mapping the data into a higher dimensional space using polynomial functions. It can capture more complex relationships between data points.