Machine Learning

Brief Answers

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 compared to RSS. R-squared represents the proportion of the variance in the dependent variable that is predictable from the independent variables. It ranges from 0 to 1, where 1 indicates a perfect fit. RSS, on the other hand, measures the discrepancy between the observed and predicted values of the dependent variable. While RSS provides information about the overall fit of the model, R-squared gives an indication of how much of the variance in the dependent variable is explained by the independent variables.

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) represents the total variance in the dependent variable, Explained Sum of Squares (ESS) represents the variance explained by the regression model, and Residual Sum of Squares (RSS) represents the unexplained variance or the discrepancy between the observed and predicted values. The equation relating these three metrics is: TSS = ESS + RSS.

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

   The need for regularization in machine learning arises to prevent overfitting and improve the generalization of the model. Regularization techniques introduce a penalty term to the loss function, discouraging overly complex models that might fit the training data too closely and perform poorly on unseen data.

4. What is Gini–impurity index?

   The Gini impurity index is a measure of how often a randomly chosen element from the set would be incorrectly labeled if it was randomly labeled according to the distribution of labels in the subset. It is

commonly used in decision tree algorithms, such as CART (Classification and Regression Trees), to evaluate the impurity of a node.

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

   Yes, unregularized decision trees can be prone to overfitting, especially when they are deep and complex. This is because decision trees can learn intricate patterns in the training data, including noise, which may not generalize well to unseen data.

6. What is an ensemble technique in machine learning?

   Ensemble techniques in machine learning involve combining the predictions of multiple individual models to improve overall performance. This can lead to better generalization, robustness, and accuracy compared to using a single model.

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

   Bagging involves training multiple instances of the same base learning algorithm on different subsets of the training data, with replacement. Predictions are then averaged or aggregated to make the final prediction. Boosting, on the other hand, sequentially trains multiple weak learners, with each subsequent learner focusing more on the examples that the previous ones misclassified. The final prediction is typically a weighted sum of the individual learner predictions.

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

   In random forests, the out-of-bag (OOB) error is the error rate of the model on the training data points that were not included in the bootstrap sample used to train each individual tree. It provides an estimate of how well the model will generalize to unseen data 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. The training data is divided into k subsets, or folds, and the model is trained k times, each time using k-1 folds for training and the remaining fold for

validation. The performance metric is then averaged over the k iterations to obtain a more reliable estimate of model performance.

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

Hyperparameter tuning involves optimizing the hyperparameters of a machine learning algorithm to improve its performance. This is done by searching through a predefined hyperparameter space using techniques such as grid search or randomized search. Hyperparameter tuning is necessary because the choice of hyperparameters can significantly impact the performance of the model.

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

If the learning rate in gradient descent is too large, it can lead to issues such as divergence, where the model fails to converge to the optimal solution, or oscillation, where the model's parameters fluctuate around the optimal solution without converging. This can result in slow convergence or even failure to converge altogether.

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

Logistic Regression is a linear classifier and may not perform well on nonlinear data since it can only capture linear decision boundaries. For nonlinear data, more complex models like Support Vector Machines (SVMs) or neural networks are typically more suitable.

13. Differentiate between Adaboost and Gradient Boosting.

Adaboost and Gradient Boosting are both ensemble learning techniques that sequentially combine multiple weak learners to create a strong learner. However, they differ in how they assign weights to training instances and update the model. Adaboost focuses on instances that are difficult to classify, while Gradient

Boosting fits each subsequent model to the residual errors of the previous model.

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

The bias-variance trade-off is a fundamental concept in machine learning that describes the trade-off between bias and variance in the performance of a model. Bias refers to the error introduced by approximating a real-world problem with a simplified model, while variance refers to the model's sensitivity to small fluctuations in the training data. A model with high bias tends to underfit the data, while a model with high variance tends to overfit the data.

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

The linear kernel is the simplest kernel function, and it represents a linear decision boundary. It is suitable for linearly separable data.

The RBF kernel is a popular choice for SVMs because it can capture complex nonlinear relationships in the data. It is defined as exp(-gamma * ||x - x'||^2), where gamma is a hyperparameter that controls the kernel's smoothness.

The polynomial kernel allows SVMs to capture polynomial decision boundaries of varying degrees. It is defined as (gamma * (x . x' + r)^d), where gamma, r, and d are hyperparameters controlling the kernel's behavior.