MACHINE LEARNING

ASSIGNMENT - 5

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?

ANS:- A higher R-squared value indicates a higher amount of variability being explained by our model and vice-versa. If we had a really low RSS value, it would mean that the regression line was very close to the actual points. This means the independent variables explain the majority of variation in the target variable

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.

ANS:- The total sum of squares (TSS) measures how much variation there is in the observed data, while the RSS(residual sum of squares) measures the variation in the error between the observed data and modeled values.The explained sum of squares (ESS) is the sum of the squares of the deviations of the predicted values from the mean value of a response variable, in a standard regression model

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

ANS:- Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting.

4. What is Gini–impurity index?

ANS:- Gini Index, also known as Gini impurity, calculates the amount of probability of a specific feature that is classified incorrectly when selected randomly. If all the elements are linked with a single class then it can be called pure.

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

ANS:- Overfit condition arises when the model memorizes the noise of the training data and fails to capture important patterns. A perfectly fit decision tree performs well for training data but performs poorly for unseen test data

Decision trees are prone to overfitting, especially when a tree is particularly deep. This is due to the amount of specificity we look at leading to smaller sample of events that meet the previous assumptions. This small sample could lead to unsound conclusions.

6. What is an ensemble technique in machine learning?

ANS:- Ensemble methods are techniques that create multiple models and then combine them to produce improved results. Ensemble methods usually produces more accurate solutions than a single model would. This has been the case in a number of machine learning competitions, where the winning solutions used ensemble methods.

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

Bagging	Boosting
Objective to decrease variance, not bias.	Objective to decrease bias, not variance.
Each model is built independently.	New models are affected by the implementation of the formerly developed model.
It is the simplest way of connecting predictions that belong to a similar type.	It is a method of connecting predictions that belong to multiple types.
Bagging tries to tackle the over-fitting problem.	Boosting tries to reduce bias.
Several training data subsets are randomly drawn with replacement from the whole training dataset.	Each new subset includes the components that were misclassified by previous models.
Bagging can solve the over-fitting problem.	Boosting can boost the over-fitting problem.

8. What is out-of-bag error in random forests?
ANS:- Out-of-bag (OOB) error, also called out-of-bag estimate, is a method of measuring the prediction error of random forests, boosted decision trees, and other machine learning models utilizing bootstrap aggregating (bagging). Bagging uses subsampling with replacement to create training samples for the model to learn from.

9. What is K-fold cross-validation?
ANS:- Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample. The procedure has a single parameter called k that refers to the number of groups that a given data sample is to be split into. As such, the procedure is often called k-fold cross-validation

10. What is hyper parameter tuning in machine learning and why it is done?
ANS Hyperparameter tuning consists of finding a set of optimal hyperparameter values for a learning algorithm while applying this optimized algorithm to any data set. That combination of hyperparameters maximizes the model's performance, minimizing a predefined loss function to produce better results with fewer errors.

11. What issues can occur if we have a large learning rate in Gradient Descent?
ANS:- In order for Gradient Descent to work, we must set the learning rate to an appropriate value. This parameter determines how fast or slow we will move towards the optimal weightsA learning rate that is too large can cause the model to converge too quickly to a suboptimal solution, whereas a learning rate that is too small can cause the process to get stuck.

12. Can we use Logistic Regression for classification of Non-Linear Data? If not, why?
ANS:- No,we cannot use Logistic Regression for classification of Non-Linear Data. It can only be used to predict discrete functions. Hence, the dependent variable of Logistic Regression is bound to the discrete number set. It is very fast at classifying unknown records. Non-linear problems can't be solved with logistic regression because it has a linear decision surface

13. Differentiate between Adaboost and Gradient Boosting.
ANS:-

Adaboost	Gradient Boosting
An additive model where shortcomings of previous models are identified by high-weight data points.	An additive model where shortcomings of previous models are identified by the gradient.
The trees are usually grown as decision stumps	The trees are grown to a greater depth usually ranging from 8 to 32 terminal nodes.
Each classifier has different weights assigned to the final prediction based on its performance	All classifiers are weighed equally and their predictive capacity is restricted with learning rate to increase accuracy.
It gives weights to both classifiers and observations thus capturing maximum variance within data.	It builds trees on previous classifier’s residuals thus capturing variance in data.

14. What is bias-variance trade off in machine learning?
ANS:- In statistics and machine learning, the bias–variance tradeoff is the property of a model that the variance of the parameter estimated across samples can be reduced by increasing the bias in the estimated parameters.

15. Give short description each of Linear, RBF, Polynomial kernels used in SVM.
ANS:-Linear Kernel is used when the data is Linearly separable, that is, it can be separated using a single Line. It is one of the most common kernels to be used. It is mostly used when there are a Large number of Features in a particular Data Set.

RBF is the default kernel used within the sklearn's SVM classification algorithm and can be described with the following formula: where gamma can be set manually and has to be >0.

In machine learning, the polynomial kernel is a kernel function commonly used with support vector machines (SVMs) and other kernelized models, that represents the similarity of vectors (training samples) in a feature space over polynomials of the original variables, allowing learning of non-linear models.