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?

R - squared is the good fit because of the following reason:

o R-squared is a statistical method that determines the goodness of fit.

o It measures the strength of the relationship between the dependent and independent variables on a scale of 0-100%.

o The high value of R-square determines the less difference between the predicted values and actual values and hence represents a good model.

o It is also called a coefficient of determination, or coefficient of multiple determination for multiple regression.

o It can be calculated from the below formula:

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 is the squared differences between the observed dependent variable and its mean. You can think of this as the dispersion of the observed variables around the mean – much like the variance in descriptive statistics.

Residual Sum of Squares (RSS) is a statistical method that helps identify the level of discrepancy in a dataset not predicted by a regression model. Thus, it measures the variance in the value of the observed data when compared to its predicted value as per the regression model.

The Explained Sum of Squares (ESS) tells you how much of the variation in the dependent variable your model explained. Explained SS = Σ(Y-Hat – mean of Y)2.

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

--> Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting. Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.

4. What is Gini–impurity index?

--> The Gini Index or Gini Impurity is calculated by subtracting the sum of the squared probabilities of each class from one. It favours mostly the larger partitions and are very simple to implement. In simple terms, it calculates the probability of a certain randomly selected feature that was classified incorrectly.

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

--> 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?

--> Ensemble methods are techniques that aim at improving the accuracy of results in models by combining multiple models instead of using a single model. The combined models increase the accuracy of the results significantly. This has boosted the popularity of ensemble methods in machine learning.

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

--> Bagging is a technique for reducing prediction variance by producing additional data for training from a dataset by combining repetitions with combinations to create multi-sets of the original data. Boosting is an iterative strategy for adjusting an observation's weight based on the previous classification.

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

--> Out of bag error in the random forests is used to evaluate the random forest model. As we know in random forest a number of decision trees are trained in parallel on bootstrapped samples. While training a particular tree the data points which are not used for training of that tree act as unseen data.For each data point predictions are made by those trees in whose training these points were not included and error on these predictions are called out of bag error.

9. What is K-fold cross-validation?

→ K-fold cross validation is the model evaluation technique which is generally used when we have limited data. In this technique we create k groups on the training data, train the model on k-1 groups and test the resultant model on the left-out group. In this we do it on every possible combinations ofgroups and then take average of the evaluation metric.

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

→ Hyper parameter tuning is the technique of finding best possible values of a set of hyper parameters used in model on which the model gives best performance. So, hyper parameter tuning is like fine tuning your model on the basis of hyper parameter values used in the model so that we get the best possible version of that model.

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

→ The two main issues which can occur if we perform gradient descent with large learning rate are:

• The gradient descent algorithm can diverge from the optimal solution if we try out a very large learning rate. The algorithm can go away from the optimal solution if we have a very large learning rate.

• The gradient may simply keep oscillating around the optimal solution if the learning rate is high, and it will not settle at the optimal solution.

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

→ We cannot use Logistic Regression for classification of Non-Linear Data because the decision boundary produced by logistic regression is linear and if we have nonlinear data where we have nonlinear decision boundaries then if we try to use the logistic regression it will perform poor on the data, as the decision boundary in the data is nonlinear.

13. Differentiate between Adaboost and Gradient Boosting.

→ Adaboost and Gradient Boosting are ensemble techniques, in which the trees are trained in series. The major difference between Adaboost and Gradient Boosting is that in Adaboost we assign weights to each of the data points of the training data and the weights changes according to the errors made by the previous tree in the series. So basically a tree in Adaboost puts more emphasis on the data-points on which the previous tree did not perform well.

While in Gradient Boosting, each tree is trained on the errors made by the previous tree, so in this way the error made on the training data keeps on decreasing as we keep increasing the trees in the series.

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

→ The bias variance tradeoff is the tradeoff which happens between bias (error made by the model) and variance (how much the model changes with change in training data) when the model complexity changes. If the model is too simple the model makes too much errors and its predictions becomes inaccurate, so when we decrease the model complexity the bias (or errors) increases but the variance decreases (that is the model does not change much with change in training data). On the other hand, if model is too complex the bias although becomes low but the variance increases. So, there is tradeoff between bias and variance. So, we always have to find that sweet spot where the model does not have much bias neither much high variance.

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

→ The SVM uses kernel functions to transform the data from one set of dimensions to another set of dimensions so that the decision boundary in the resultant space is simpler than the decision boundary in original space. Now, the kernel to be used depend upon the nature of the data we have.

- Linear kernel will be used if the original data is linearly separable.
- Polynomial kernel is used when the data is the form of polynomial of some degree n.
- RBF is used when the data follows some complex pattern which is neither linear nor polynomial.