MACHINE LEARNING

FLIP ROBO

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: R-squared (R^2) and Residual Sum of Squares (RSS) are both commonly used measures to assess the goodness of fit of a regression model, but they capture different aspects of model performance, and the choice between them depends on the context and what you want to evaluate.

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: In particular, the explained sum of squares measures how much variation there is in the modelled values and this is compared to the total sum of squares (TSS), which measures how much variation there is in the observed data, and to the residual sum of squares, which measures the variation in the error between the ...

formula.

$SST=n\sum i=1(yi-\bar{}y)2\square\square\square=\sum\square=1\square(\square\square-\square\bar{})2$

Where:

$yi\ \square\square$ – observed dependent variable

$\bar{}y\ \square\bar{}$ – mean of the dependent variable

The SSR formula is the following:

$SSR=n\sum i=1(\hat{}yi-\bar{}y)2\square\square\square=\sum\square=1\square(\square\hat{}\square-\square\bar{})2$

Where:

$\hat{}yi\ \square\hat{}\square$ – the predicted value of the dependent variable

$\bar{}y\ \square\bar{}$ – mean of the dependent variable

If **SSR** equals **SST**, our **regression model** perfectly captures all the observed variability, but that's rarely the case.

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

ANS: While training a machine learning model, the model can easily be overfitted or under fitted. To avoid this, we use regularization in machine learning to properly fit a model onto

our test set. Regularization techniques help reduce the chance of overfitting and help us get an optimal model.

2. What is Gini–impurity index?
   ANS: Gini Impurity tells us what is the probability of misclassifying an observation. Note that the lower the Gini the better the split. In other words the lower the likelihood of misclassification.

5. Are unregularized decision-trees prone to overfitting? If yes, why?
   ANS: Decision trees are prone to overfitting when they capture noise in the data. Pruning and setting appropriate stopping criteria are used to address this assumption.

3. 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 in machine learning usually produce more accurate solutions than a single model would.

4. What is the difference between Bagging and Boosting techniques?
   ANS:
   Bagging and boosting are different ensemble techniques that use multiple models to reduce error and optimize the model. The bagging technique combines multiple models trained on different subsets of data, whereas boosting trains the model sequentially, focusing on the error made by the previous model.

5. What is out-of-bag error in random forests?
   ANS: The out-of-bag (OOB) error is the average error for each calculated using predictions from the trees that do not contain in their respective bootstrap sample. This allows the RandomForestClassifier to be fit and validated whilst being trained.

6. What is K-fold cross-validation?
   ANS: K-fold cross-validation is a technique for evaluating predictive models. The dataset is divided into k subsets or folds. The model is trained and evaluated k times, using a different fold as the validation set each time. Performance metrics from each fold are averaged to estimate the model's generalization performance.

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

8. What issues can occur if we have a large learning rate in Gradient Descent?
   ANS: If the learning rate is too high, the algorithm may overshoot the minimum, and if it is too low, the algorithm may take too long to converge. Over fitting: Gradient descent can over fit the training data if the model is too complex or the learning rate is too high.

9. Can we use Logistic Regression for classification of Non-Linear Data? If not, why?
   ANS:
   Logistic Regression does not assume linear relationship between dependent and independent variables as it applies a non linear log transformation. But there could be some linear relationship among the independent variables i.e. multiCollinearity which anyways should be avoided, in general, in any model.

10. Differentiate between Adaboost and Gradient Boosting.

ANS: Comparing Gradient Boosting and Ada Boost

Then Ada Boost builds a new stump based on the errors that the previous stump made. Ada Boost continues to make stumps in this fashion until it has made the number of stumps we've asked for. In Contrast, Gradient Boost starts by making a single leaf instead of a tree or stump.

11. What is bias-variance trade off in machine learning?
    ANS: In statistics and machine learning, the bias–variance tradeoff describes the relationship between a model's complexity, the accuracy of its predictions, and how well it can make predictions on previously unseen data that were not used to train the model.

12. Give short description each of Linear, RBF, Polynomial kernels used in SVM.
    ANS: **RBF Kernel is popular because of its similarity to K-Nearest Neighborhood Algorithm. It has the advantages of K-NN and overcomes the space complexity problem as RBF Kernel Support Vector Machines just needs to store the support vectors during training and not the entire dataset.**
    An SVM with a linear kernel learns a linear decision boundary in the original feature space. A kernel SVM, on the other hand still learns a linear decision boundary, but in a transformed space. For example, a radial basis function SVM learns a linear boundary in an infinite dimensional space.