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 smaller or lower value for the residual sum of squares is ideal in any model since it means there's less variation in the data set. In other words, the lower the sum of squared residuals, the better the regression model is at explaining the data.

Adding more independent variables or predictors to a regression model tends to increase the R-squared value, which tempts makers of the model to add even more. Adjusted R-squared is used to determine how reliable the correlation is and how much is determined by the addition of independent variables.

1. 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 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](https://en.wikipedia.org/wiki/Regression_model) — for example, yi = a + b1x1i + b2x2i + ... + εi, where yi is the i th observation of the [response variable](https://en.wikipedia.org/wiki/Response_variable), xji is the i th observation of the j th [explanatory variable](https://en.wikipedia.org/wiki/Explanatory_variable), a and bj are [coefficients](https://en.wikipedia.org/wiki/Coefficient), i indexes the observations from 1 to n, and εi is the i th value of the [error term](https://en.wikipedia.org/wiki/Error_term). In general, the greater the ESS, the better the estimated model performs.

TSS = ESS + RSS, where TSS is Total Sum of Squares, ESS is Explained Sum of Squares and RSS is Residual Sum of Suqares. The aim of Regression Analysis is explain the variation of dependent variable Y.

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

Ans= This is a form of regression, that constrains/ regularizes or shrinks the coefficient estimates towards zero. In other words, this technique discourages learning a more complex or flexible model, so as to avoid the risk of overfitting.

In general, regularization means to make things regular or acceptable. In the context of machine learning, regularization is the process which regularizes or shrinks the coefficients towards zero. In simple words, regularization discourages learning a more complex or flexible model, to prevent overfitting.

1. What is Gini–impurity index?

Ans=Gini index or Gini impurity measures the degree or probability of a particular variable being wrongly classified when it is randomly chosen.

If all the elements belong to a single class, then it can be called pure. The degree of Gini index varies between 0 and 1, where 0 denotes that all elements belong to a certain class or if there exists only one class, and 1 denotes that the elements are randomly distributed across various classes. A Gini Index of 0.5 denotes equally distributed elements into some classes.

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

Ans=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.

The common pattern for overfitting can be seen on learning curve plots, where model performance on the training dataset continues to improve.

1. 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.

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

Ans=Bagging is a method of merging the same type of predictions. Boosting is a method of merging different types of predictions. Bagging decreases variance, not bias, and solves over-fitting issues in a model. Boosting decreases bias, not variance.

1. 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.

1. What is K-fold cross-validation?

Ans=Cross-validation is a statistical method used to estimate the skill of machine learning models.It is commonly used in applied machine learning to compare and select a model for a given predictive modeling problem because it is easy to understand, easy to implement, and results in skill estimates that generally have a lower bias than other methods.

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

Ans=Hyperparameters tuning is crucial as they control the overall behavior of a machine learning model. Every machine learning models will have different hyperparameters that can be set. A hyperparameter is a parameter whose value is set before the learning process begins.

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

Ans=When the learning rate is too large, gradient descent can inadvertently increase rather than decrease the training error.When the learning rate is too small, training is not only slower, but may become permanently stuck with a high training error.

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

Ans=Logistic Regression has traditionally been used as a linear classifier, i.e. when the classes can be separated in the feature space by linear boundaries. That can be remedied however if we happen to have a better idea as to the shape of the decision boundary.

Logistic regression is a powerful machine learning algorithm that utilizes a sigmoid function and works best on binary classification problems, although it can be used on multi-class classification problems through the “one vs. all” method. Logistic regression (despite its name) is not fit for regression tasks.

13. Differentiate between Adaboost and Gradient Boosting.

Ans=The main differences, therefore, are that Gradient Boosting is a generic algorithm to find approximate solutions to the additive modeling problem, while AdaBoost can be seen as a special case with a particular loss function.In Adaboost, 'shortcomings' are identified by high-weight data points.

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

Ans=Bias is the simplifying assumptions made by the model to make the target function easier to approximate. Variance is the amount that the estimate of the target function will change given different training data. Trade-off is tension between the error introduced by the bias and the variance.

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

Ans=SVM Kernel Functions SVM algorithms use a set of mathematical functions that are defined as the kernel. The function of kernel is to take data as input and transform it into the required form. These functions can be different types. For example linear, nonlinear, polynomial, radial basis function (RBF), and sigmoid.