Machine Learning Assignment -5

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 →

Both R-squared and Residual Sum of Squares (RSS) are measures of goodness of fit in regression analysis, but they capture different aspects of the model's performance.

R-squared indicates how well the model fits the data, with values ranging from 0 to 1. Higher R-squared values indicate a better fit, as they mean that a larger proportion of the variation in the dependent variable is explained by the independent variables in the model.

On the other hand, RSS measures the total sum of squared differences between the actual values of the dependent variable and the predicted values by the model. It represents the amount of unexplained variation in the data, and lower RSS values indicate a better fit.

Therefore, both measures are useful in evaluating the goodness of fit of a model, but they serve different purposes. R-squared is a useful measure to assess the overall fit of the model and to compare different models, while RSS is useful to identify the degree of the error in the model's predictions.

In general, a good model should have both a high R-squared value and a low RSS value, indicating that it explains a large proportion of the variation in the dependent variable and has a low degree of error in its predictions. However, in some cases, one measure may be more important than the other, depending on the research question and the nature of the data being analysed.

both measures can be valuable, Goodness of fit is best explained for R-squared, and RSS are best for finding model error rather than fit. The choice between them should align with your specific goals and the questions you want to answer about your regression model.

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 SS (TSS or SST) tells you how much variation there is in the dependent variable.
The Explained SS tells you how much of the variation in the dependent variable your model explained.
The residual sum of squares tells you how much of the dependent variable's variation your model did not explain.

$$TSS = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

TSS = Total sum of squares
n = number of observations
y_i = value in a sample
$\bar{y}$ = mean value of sample
TSS = ESS + RSS

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

Ans →

Regularization refers to techniques that are used to calibrate machine learning models to minimize the adjusted loss function and prevent overfitting or underfitting.

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.

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?

Ans →

For a single sample that is taken from the data set, the likelihood of being misclassified if it were given a random class label describes the Gini impurity.

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

Ans →

Yes, in decision trees, to fit the data, the model keeps generating new nodes and ultimately the tree becomes too complex to interpret. The decision tree predicts well for the training data but can be inaccurate for new data. If a decision tree model is allowed to train to its full potential, it can overfit the training data.

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

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

Ans →

Bagging	Boosting
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.
Each model receives an equal weight.	Models are weighed based on their performance.
Models are built independently in Bagging.	New models are affected by a previously built model's performance in Boosting.
Training data subsets are drawn randomly with a replacement for the training dataset.	Every new subset comprises the elements that were misclassified by previous models.
Bagging is usually applied where the classifier is unstable and has a high variance.	Boosting is usually applied where the classifier is stable and simple and has high bias.

8. 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.

9. 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.

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

Ans →

Hyperparameter tuning is an essential part of controlling the behaviour of a machine learning model. If we don't correctly tune our hyperparameters, our estimated model parameters produce suboptimal results, as they don't minimize the loss function. This means our model makes more errors.

11. 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.
Overfitting: Gradient descent can overfit the training data if the model is too complex or the learning rate is too high.

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

Ans →

Logistic regression has traditionally been used to come up with a hyperplane that separates the feature space into classes. But if we suspect that the decision boundary is nonlinear, we may get better results by attempting some nonlinear functional forms for the logit function.

13. Differentiate between Adaboost and Gradient Boosting.

Ans →

Adaboost	Gradient Boost
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 →

The bias–variance trade off 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.

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

Ans →

The linear kernel produces a decision boundary that is a hyperplane in the feature space. This hyperplane separates data points from different classes in a linear fashion. It assumes that the relationship between the features and the target variable is linear.

The radial basis function kernel is a popular kernel function used in various kernelized learning algorithms. It is commonly used in support vector machine classification.

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