MACHINE LEARNING WORKSHEET-5

1. R-squared is a better measure of goodness of fit in regression because it measures the proportion of variance in the dependent variable that is explained by the independent variables. On the other hand, RSS measures the amount of error in the model, which is not necessarily related to how well the model fits the data. R-squared also provides a more intuitive understanding of the strength of the model, since it is expressed as a percentage.
2. TSS (Total Sum of Squares): Total sum of squares is the sum of squared differences between the observed (actual) responses and their mean. It is a measure of the total variance in the response variable.
   ESS (Explained Sum of Squares): Explained sum of squares is the sum of squared differences between the predicted responses and their mean. It is a measure of the variance that is explained by the regression model.
   RSS (Residual Sum of Squares): Residual sum of squares is the sum of squared differences between the observed (actual) responses and their predicted responses. It is a measure of the variance that is not explained by the regression model.
   The relationship between TSS, ESS and RSS can be expressed as follows:
   TSS = ESS + RSS
3. Regularization is a technique used in machine learning to prevent overfitting. It adds a penalty to the loss function to reduce the complexity of the model and prevent it from overfitting on the training data. This helps to improve the accuracy of the model on unseen data. Regularization also helps to improve the generalization ability of the model, making it more robust in the presence of noisy data.
4. The Gini–impurity index is a measure of the probability of misclassifying a randomly chosen element in a dataset, if it were randomly labeled according to the class distribution in the dataset. It is used as a measure of the impurity of an arbitrary collection of examples. It is also used in decision tree learning and in random forests.
5. Yes, unregularized decision trees are prone to overfitting. This is because unregularized decision trees have the ability to create very complex models that can fit the training data very well, but that may not generalize well to unseen data. The tree can become too complex and will end up memorizing the training data, rather than learning the underlying pattern.
6. Ensemble techniques in machine learning are methods that combine multiple machine learning algorithms to obtain better predictive performance than could be obtained from any of the individual models alone. Ensemble techniques are used to improve the stability and predictive power of models by combining multiple weak learners to form a strong learner. Examples of ensemble techniques include bagging, boosting, and stacking.
7. Bagging (also known as Bootstrap Aggregation) is an ensemble method that combines multiple weaker models to create a stronger model. It works by training each model on a separate subset of the data and then combining the predictions of each model to make a more accurate final prediction.
   Boosting is an ensemble method that combines several weak models to create a stronger model. It works by training each model sequentially, with each model focusing on the errors made by the previous model. The output of each model is then combined to create a more accurate final prediction. Unlike bagging, boosting pays more attention to the data points that the previous model had difficulty with, thus allowing for a more accurate final prediction.
8. Out-of-bag error in random forests is the error rate on the data that is not used in the training set for each tree. During the training process, some of the data is left out of the bootstrap sample for each tree. This data is then used to test the generalization of the model, and the out-of-bag error is the average error rate on this data. This is a measure of the generalization performance of the random forest model.
9. K-fold cross-validation is a technique used in machine learning to evaluate the accuracy of a model on unseen data. It involves randomly splitting a dataset into k-folds (or parts) of equal

size. The model is then trained on k-1 folds, and tested on the remaining fold. The performance metric is then calculated using the results from the k-folds. This process is repeated k times, and the average of the k-fold performance metric is taken as the final result. This technique helps to reduce bias in the model, as every data point is used in both the training and test set.

10. Hyperparameter tuning is the process of selecting the optimal values for a model's hyperparameters in order to maximize its performance on unseen data. It is done to optimize the model’s performance, increase predictive accuracy, and reduce overfitting.
11. If the learning rate is too large, the model has a higher chance of overshooting the minimum of the loss function, leading to oscillations and slow convergence. Additionally, the model may not converge and may diverge, resulting in a non-optimal solution or even no solution at all.
12. No, we cannot use logistic regression for classification of non-linear data. Logistic regression is designed to work with linear data. Non-linear data requires more complex algorithms such as support vector machines and neural networks to be able to accurately classify them.
13. Adaboost and Gradient Boosting are both ensemble methods used in machine learning to create a strong predictor from multiple weak predictors. Adaboost is an ensemble technique that uses a set of weak learners which iteratively learn from the mistakes of the previous weak learners and in the end produce a strong predictor. Gradient Boosting is an ensemble technique that combines the predictions of multiple weak learners to build a strong predictor. The weak learners in Gradient Boosting are decision trees, which are trained using the gradient descent optimization algorithm. Adaboost focuses on re-weighting observations, while Gradient Boosting focuses on adding weak learners to the ensemble. Adaboost is more sensitive to outliers, while Gradient Boosting is more robust to outliers. Adaboost is faster to train and predict, while Gradient Boosting can take longer to train and predict.
14. Bias-variance trade off is a balancing act between having a model that is too simplistic (high bias) or too complex (high variance). High bias models are typically underfitting the data, meaning that they are unable to accurately capture the underlying relationships in the data and therefore have low accuracy. On the other hand, high variance models are overfitting the data and are more sensitive to noise, resulting in low accuracy and poor generalization.
15. Linear Kernel: The linear kernel is the simplest and most commonly used kernel in Support Vector Machines. It is used to separate linearly separable data. It takes the form of a dot product between two vectors and outputs a single real number. RBF (Radial Basis Function) Kernel: The RBF kernel is a non-linear kernel used for data that is not linearly separable. It is frequently used for classification problems and works by mapping data into a higher-dimensional feature space, where the data may become linearly separable. Polynomial Kernel: The polynomial kernel is a non-linear kernel used to separate data that is not linearly separable. It works by transforming the data into a higher-dimensional space, where the data may become linearly separable. The kernel takes the form of a polynomial of degree n, where n is an integer value greater than one.