

## Assignment 5 Answers

1. R-Squared is better measure for goodness of fit model in regression because it is easy to interpret results as it gives results on scale of 0 to 1 whereas Residual Sum of squares gives absolute number ie., if sum of squares is higher, then it means higher variance from the mean value and vice versa that depends on scale of dependent variable therefore less interpretable without context.
2. **TSS** shows the total variation in the observed data and it is sum of squares of difference between each observation and overall mean.  
**ESS** shows the variation explained in regression model and it is sum of squares of difference between predicted values and overall mean.  
**RSS** shows the variation that is not explained by regression model and it is sum of squares of the residuals, which are the difference between observed and predicted values.

The relationship between these three metrics in equation as follows:

$$TSS=ESS+RSS$$

3. Regularization technique is used to improve the performance of models not only on training data but also on new and unseen data. Here with this technique, we can prevent overfitting of data by adding a penalty term to the model's objective function during training which leads to simpler models that are less likely to overfit and perform well on unseen data.
4. Gini impurity index is a measure to know how a randomly chosen element of set will be incorrectly labeled if it were labeled randomly and independently according to the distribution of labels in the set.
5. Yes, as the unregularized decision trees models are prone to overfitting, high variance and lack of generalization which means a tree that is too complex might achieve a perfect accuracy score on training data but will perform poorly on new and unseen data as it has high variance and predictions can be unstable.
6. An ensemble technique enhances accuracy of predictions by combining multiple learning models instead of single model to improve performance of the overall learning.
7. In Bagging predictions are combined of same type and its aim is to reduce variance. This model tries to solve overfitting of data. If the classifier is unstable (high variance), we will be using Bagging technique.  
In Boosting technique predictions of different types are combined. And its aim is to decrease bias as every new subset contains the data that were misclassified by previous model. If the classifier is stable and simple (high bias), we will use Boosting technique.
8. On unseen data, Out of Bag errors are an estimate of the performance of a random forest classifier or regressor. The error is average error rate across all trees in the forest. It is computed using samples that were not included in training of individual trees. The Out of Bag error will stabilize with high number of iterations.

9. K-fold cross-validation – In order to assess the model's ability when new data becomes available, the data set is divided into a number of K-folds. Here K represents the number of groups into which the data sample is divided. This model ensures that it generalizes well on an independent dataset.
10. Hyperparameters are configurable variables that control learning process of machine learning model. In order to increase the model's accuracy, generalization and other metrics it is used to tune the performance of the model. Hyperparameter tuning is the process of selecting the optimal values for machine learning model's hyperparameters.
11. For Gradient descent if the learning rate is high, it can suffer from divergence which means the weights increase exponentially. The Algorithm will result in exploding gradients causing problems like high loss values, numerical instabilities and it may oscillate around the minimum and not settle down due to large updates which ultimately causes unpredictable training behavior.
12. On Non linear data, as the relationship between features and target variable is not linear Logistic regression method may not perform well. For it to perform well it need linear data. However, there are other ways like by creating new feature that capture non linearity and establishes relationship between variables in a nonlinear manner we can use this method.
13. AdaBoost and Gradient Boosting are boosting techniques and they differ in their approach.  
AdaBoost uses misclassified points by the previous classifier and increasing their weights. It uses weak models and the final prediction is a weighted sum of the weak learners' predictions.  
Gradient Boosting works by improving a loss function, by adding a new model that reduces the residual errors made by previous models. It can use more complex models as weak learners.
14. In any model one will generally choose the ML technique which has low bias and low variance. Finding the correct balance between bias and variance where total error is minimized is called trade-off.
15. Under SVM Linear Kernel is easy to understand and fast as it is used when the data is linearly separable. It computes the dot product between the feature vector.  
RBF(Radial basis function) is used with handling non-linear decision boundaries. It maps the data into an infinite dimensional space. This model is used with more complex non linear kernels.  
Polynomial Kernal is effectively used for non linear data. It computes the similarity between two vectors in terms of polynomial of the original variables. This model is used for images and used to figure out what objects are in pictures by looking at the details.