# MACHINE LEARNING

**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?

   According to me, the R-squareed metric is a better measure of goodness of fit model because it is explains well the variation in the data and compares model to select the better model and it gives detailed understanding of the models accuracy.

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.

   TSS shows us how much the actual data points vary from the mean value of the dependent variable.
   ESS tells us how much of the total variation in the dependent variable is explained by our regression model.
   RSS measures the amount of variation that our model couldn't explain.
   TSS = ESS + RSS

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

   Regularisation in ML prevents overfitting of the data and helps in generalization of the data. Thus it increases the performance of the model.

4. What is Gini–impurity index?

   It measure how random the dataset is.

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

   Yes, because when we train on the unregularized data, it learns the training data only but not the underlying pattern. Thus the generalization is reduced and overfitting occurs.

6. What is an ensemble technique in machine learning?

   Ensemble techniques means combining n models to create a strong individual and robust model

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

   In bagging, the training dataset is fed into different models parallely and the models are trained parallely and the prediction from each model is averaged to get final output

   In boosting the training dataset is fed into different model sequentially and the feedback is received from every model and try to improve the results in the subsequent model.

8. What is out-of-bag error in random forests?

   The out-of-bag error is the error rate of the model calculated using only those data points that were not included in the training set of a particular tree.

9. What is K-fold cross-validation?

   K-fold cross-validation is a technique used to evaluate the performance of a machine learning model by splitting the dataset into K equal-sized subsets in which one of the subset is used as test data and the remaining as training data subset.

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

    Hyperparameter is the tuning of the parameters to train the model so that its accuracy is increased and prevent over/under fitting and improve generalization

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

    It leads to poor generalisation, convergence difficulty, unstable and skipping the minimum

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

    Logistic regression is a linear classification algorithm. It may lead to poor performance and inaccurate predictions.

13. Differentiate between Adaboost and Gradient Boosting.

    Adaboost optimizes by adjusting instance weights to improve model performance and Gradient Boosting optimizes by directly minimizing a loss function with gradient descent.

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

    It describes the tradeoff that occurs when trying to minimize both the bias and variance of a model simultaneously.

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

    Linear kernel:
    It compiutes the dot products between feature vectors in the original feature spaces
    It works well for linearly separable datasets

    RBF:
    It captures complex non-linear relationships between features. It is effective for datasets with non-linear decision boundaries

    Polynomial:
    It's helpful for sorting things that have a more complicated relationship