**Machine Learning Assignment**

1. D) All of the above: GridSearchCV(), RandomizedCV(), and K-fold Cross Validation are used for hyper parameter tuning in the sk-learn library.
2. A) Random forest: In Random forest, trees are trained in parallel.
3. A) The regularization will increase: If we increase the C hyper parameter in the line of code, regularization will increase.
4. A) It regularizes the decision tree by limiting the maximum depth up to which a tree can be grown: The max\_depth hyper parameter in the code regulates the decision tree by limiting the maximum depth up to which a tree can be grown.
5. A) It's an ensemble of weak learners: Random Forest is an ensemble of weak learners.
6. A) Gradient Descent algorithm can diverge from the optimal solution: If the learning rate is very high in gradient descent, the Gradient Descent algorithm can diverge from the optimal solution.
7. B) Bias will decrease, Variance increase: As the model complexity increases, bias will decrease, and variance will increase.
8. A) model is underfitting: If the test accuracy is significantly less than the train accuracy, it indicates that the model is underfitting.
9. The Gini index and entropy of the dataset with 40% A and 60% B classes can be calculated as follows:

Gini Index = 1 - (0.4)^2 - (0.6)^2 = 0.48

Entropy = -(0.4log2(0.4) + 0.6log2(0.6)) = 0.97

1. Random Forests have the following advantages over Decision Tree:
   1. Random Forests are less prone to overfitting.
   2. They are faster to train.
   3. They can handle missing values and imbalanced datasets better.
   4. They can give feature importance.
2. Scaling all numerical features in a dataset is required to bring them to a common scale to avoid one feature dominating over the other. The two techniques used for scaling are: Min-Max Scaling and Standardization.
3. Scaling provides the following advantages in optimization using gradient descent algorithm:
   1. It helps in reaching the optimal solution faster.
   2. It makes the gradient descent algorithm less sensitive to the learning rate.
   3. It makes it easier to compare the impact of different features on the output variable.
4. No, accuracy is not a good metric to measure the performance of the model in case of a highly imbalanced dataset for a classification problem. This is because accuracy considers all classes to be equally important, which can lead to a skewed result. Instead, metrics such as F1-score, Precision, and Recall are preferred.
5. F-score is a metric that is used to evaluate the performance of a model in binary classification problems. It is the harmonic mean of precision and recall. Its mathematical formula is as follows:

F-score = 2 \* (Precision \* Recall) / (Precision + Recall)

1. In machine learning, fit() is used to train the model on the dataset, transform() is used to apply the transformation on the dataset, and fit\_transform() is used to perform both operations at once.