Name of the Assignment: Machine Learning (Worksheet-7)	
Submitted by	: Shalini Joshi
Designation	: Data Science Intern
Date of Submission	: 7th Mar,2023

1. Which of the following in sk-learn library is used for hyper parameter tuning?

A) GridSearchCV() B) RandomizedCV() C) K-fold Cross Validation D) All of the above

Ans: A) GridSearchCV()

2. In which of the below ensemble techniques trees are trained in parallel?

A) Random forest B) Adaboost C) Gradient Boosting D) All of the above

Ans: A) Random forest

3. In machine learning, if in the below line of code: sklearn.svm.SVC (C=1.0, kernel='rbf', degree=3) we increasing the C hyper parameter, what will happen?

A) The regularization will increase B) The regularization will decrease C) No effect on regularization D) kernel will be changed to linear

Ans: A) The regularization will increase

4. Check the below line of code and answer the following questions: sklearn.tree.DecisionTreeClassifier(*criterion='gini',splitter='best',max_depth=None, min_samples_split=2) . Which of the following is true regarding max_depth hyper parameter?

A) It regularizes the decision tree by limiting the maximum depth up to which a tree can be grown. B) It denotes the number of children a node can have. C) both A & B D) None of the above

Ans: A) It regularizes the decision tree by limiting the maximum depth up to which a tree can be grown.

5. Which of the following is true regarding Random Forests?

A) It's an ensemble of weak learners. B) The component trees are trained in series C) In case of classification problem, the prediction is made by taking mode of the class labels predicted by the component trees. D)None of the above

Ans: C) In case of classification problem, the prediction is made by taking mode of the class labels predicted by the component trees

6. What can be the disadvantage if the learning rate is very high in gradient descent?

A) Gradient Descent algorithm can diverge from the optimal solution. B) Gradient Descent algorithm can keep oscillating around the optimal solution and may not settle. C) Both of them D) None of them

Ans: C) Both of them

7. As the model complexity increases, what will happen?

A) Bias will increase, Variance decrease B) Bias will decrease, Variance increase C) both bias and variance increase D) Both bias and variance decrease.

Ans: B) Bias will decrease, Variance increase

8. Suppose I have a linear regression model which is performing as follows: Train accuracy=0.95 and Test accuracy=0.75 Which of the following is true regarding the model?

A) model is underfitting B) model is overfitting C) model is performing good D) None of the above

Ans: C) model is performing good

Q9 to Q15 are subjective answer type questions, Answer them briefly.

9. Suppose we have a dataset which have two classes A and B. The percentage of class A is 40% and percentage of class B is 60%. Calculate the Gini index and entropy of the dataset.

Ans: Here, p(A) = 0.40 and p(B)= 0.60

Below are the calculations for Gini Index and entropy:

Gini index = 1- (p(A)^2 + p(B)^2)

= 1-((0.4)^2+(0.6)^2)

= 1-0.52

=**0.48**

Entropy = -(p(A) log2(p(A)) + p(B) log2(p(B)))

= -(0.4*log2(0.4) + 0.6*log2(0.6))

=-(-0.03876 + 0.047509)

=-**0.008745**

10. What are the advantages of Random Forests over Decision Tree?

Ans: Briefly, although decision trees have a low bias / are non-parametric, they suffer from a high variance which makes them less useful for most practical applications. By aggregating multiple decision trees, one can reduce the variance of the model output significantly, thus improving performance.

While this could be archived by simple tree bagging, the fact that each tree is build on a bootstrap sample of the same data gives a lower bound on the variance reduction, due to correlation between the individual trees. Random Forest addresses this problem by sub-sampling features, thus de-correlating the trees to a certain extend and therefore allowing for a greater variance reduction / increase in performance.

11. What is the need of scaling all numerical features in a dataset? Name any two techniques used for scaling.

Ans: Scaling is a preprocessing method used to transform continuous data to make it look normally distributed. In scikit-learn this is often a necessary step because many models assume that the data you are training on is normally distributed, and if it isn't, your risk biasing your model.

There are two primary scaling techniques used.

- The first is standard scaling (or z-scaling) and is calculated by subtracting the mean and dividing by the standard deviation.
- The second is min-max scaling and is calculated by subtracting by the minimum value and dividing by the difference between the maximum and minimum values.

12. Write down some advantages which scaling provides in optimization using gradient descent algorithm.

Ans: Advantages which scaling provides in optimization using gradient descent algorithm:

- More stable convergence and error gradient than Stochastic Gradient descent
- Embraces the benefits of vectorization
- A more direct path is taken towards the minimum
- Computationally efficient since updates are required after the run of an epoch

13. In case of a highly imbalanced dataset for a classification problem, is accuracy a good metric to measure the performance of the model. If not, why?

Ans: Accuracy is probably not a good metric to measure the performance of the model when the dataset for classification problem, is highly imbalanced.

When the skew in the class distributions are severe, accuracy can become an unreliable measure of model performance. The reason for this unreliability is centered around the average machine learning practitioner and the intuitions for classification accuracy.

Typically, classification predictive modeling is practiced with small datasets where the class distribution is equal or very close to equal. Therefore, most practitioners develop an intuition that large accuracy score (or conversely small error rate scores) are good, and values above 90 percent are great. Achieving 90 percent classification accuracy, or even 99 percent classification accuracy, may be trivial on an imbalanced classification problem.

This means that intuitions for classification accuracy developed on balanced class distributions will be applied and will be wrong, misleading the practitioner into thinking that a model has good or even excellent performance when it, in fact, does not.

14. What is “f-score" metric? Write its mathematical formula.

Ans: The F-score, also called the F1-score, is a measure of a model’s accuracy on a dataset. It is used to evaluate binary classification systems, which classify examples into ‘positive’ or ‘negative’. The F-score is a way of combining the precision and recall of the model, and it is defined as the harmonic mean of the model’s precision and recall.

The F-score is commonly used for evaluating information retrieval systems such as search engines, and also for many kinds of machine learning models, in particular in natural language processing.

Mathematical Formula:

$$F_1 = \frac{2}{\frac{1}{\text{recall}} \times \frac{1}{\text{precision}}} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$
$$= \frac{\text{tp}}{\text{tp} + \frac{1}{2}(\text{fp} + \text{fn})}$$

15. What is the difference between fit(), transform() and fit_transform()?

Ans: **fit** means to fit the pre-processor to the data being provided. This is where the pre-processor "learns" from the data.

transform means to transform the data (produce outputs) according to the fitted pre-processor; it is normally used on the *test* data, and unseen data in general (e.g. in new data that come after deploying a model).

fit_transform means to do both - Fit the pre-processor to the data, then transform the data according to the fitted pre-processor. Calling fit_transform is a convenience to avoid needing to call fit and transform sequentially on the same input, but of course this is only applicable to the *training* data (calling again fit_transform in test or unseen data is unfortunately a common rookie mistake).

********** **The End** **********