**MACHINE LEARNING – WORKSHEET 4**

**In Q1 to Q8, only one option is correct, Choose the correct option:**

Q1. Which of the following in sklearn library is used for hyper parameter tuning?

Answer: A) GridSearchCV()

Q2. In which of the below ensemble techniques trees are trained in parallel?

Answer: A) Random forest

Q3. 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?

Answer: B) The regularization will decrease

Q4. 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?

Answer: C) both A & B

Q5. Which of the following is true regarding Random Forests?

Answer: D)None of the above

Q6. What can be the disadvantage if the learning rate is very high in gradient descent?

Answer: C) Both of them

Q7. As the model complexity increases, what will happen?

Answer: B) Bias will decrease, Variance increase

Q8. Suppose I have a linear regression model which is performing as follows:

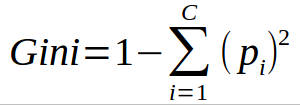
Train accuracy=0.95 Test accuracy=0.75 Which of the following is true regarding the model?

Answer: C) model is performing good

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

Q9. 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.

Answer: As we know that,

Gini Index= 1 – (P(class1) ^2 + P(class2) ^2 = 1-(0.4) ^2+(0.6) ^2 =0.804



Entropy= -0.4\*log2 (0.4) - 0.6\*log2(0.6) = - (0.159) – (- 0.133) = -0.026

Q10. What are the advantages of Random Forests over Decision Tree?

Answer: random forests are a strong modelling technique and much more robust than a single decision tree. They aggregate many decision trees to limit overfitting as well as error due to bias and therefore yield useful results.

Random forest is a technique used in modelling predictions and behaviour analysis and is built on decision trees. It contains many decision trees that represent a distinct instance of the classification of data input into the random forest. The random forest technique takes consideration of the instances individually, taking the one with the majority of votes as the selected prediction.

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

Answer: Scaling is required to rescale the data and it’s used when we want features to be compared on the same scale for our algorithm. And, when all features are in the same scale, it also helps algorithms to understand the relative relationship better.

**Normalization** is a scaling technique in which values are shifted and rescaled so that they end up ranging between 0 and 1. It is also known as Min-Max scaling. Here’s the formula for normalization:

Normalization equationNormalization equation

Here, Xmax and Xmin are the maximum and the minimum values of the feature respectively.

When the value of X is the minimum value in the column, the numerator will be 0, and hence X’ is 0

On the other hand, when the value of X is the maximum value in the column, the numerator is equal to the denominator and thus the value of X’ is 1. If the value of X is between the minimum and the maximum value, then the value of X’ is between 0 and 1.

**Standardization** is another scaling technique where the values are centered around the mean with a unit standard deviation. This means that the mean of the attribute becomes zero and the resultant distribution has a unit standard deviation.

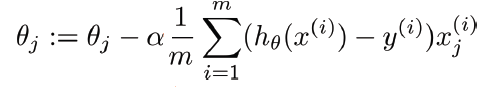
Standardization equationHere’s the formula for standardization:

Standardization equation

Feature scaling: Mu is the mean of the feature values and Feature scaling: Sigma is the standard deviation of the feature values. Note that in this case, the values are not restricted to a particular range.

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

Answer: gradient descent as an optimization technique require data to be scaled. Take a look at the formula for gradient descent below:



Gradient descent formula

feature scaling helps in causing Gradient Descent to converge much faster as standardizing all the variables on to the same scale, for example, for a linear regression makes it easy to calculate the slope ( y = mx + c) (where we normalize the M parameter to converge faster).

Q13. 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?

Answer: Classification accuracy is the number of correct predictions divided by the total number of predictions.

Accuracy can be misleading. For example, in a problem where there is a large class imbalance, a model can predict the value of the majority class for all predictions and achieve a high classification accuracy. So, further performance measures are needed such as F1 score and Brier score.

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

Answer: 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.

It is possible to adjust the F-score to give more importance to precision over recall, or vice-versa. Common adjusted F-scores are the F0.5-score and the F2-score, as well as the standard F1-score.

**F-score Formula**

The formula for the standard F1-score is the harmonic mean of the precision and recall. A perfect model has an F-score of 1.

F1 = 2 \* (precision \* recall) / (precision + recall)

In the multi-class and multi-label case, this is the average of the F1 score of each class with weighting depending on the average parameter.

Q 15. What is the difference between fit (), transform () and fit\_transform ()?

Answer: fit() - It is used for calculating the initial filling of parameters on the training data (like mean of the column values) and saves them as an internal objects state.

transform() - Use the above calculated values and return modified training data.

fit\_transform() - It joins above two steps. Internally, it just calls first fit() and then transform() on the same data.