WORKSHEET MACHINE LEARNING-7

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

➔ D) All of the above

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

➔ 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

➔ 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

➔ 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

➔ A) It's an ensemble of weak learners. 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

➔A,B

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.

➔B

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

➔B

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.
➔ Gini index = 1- (p(A)2 + p(B)2

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

) = 0.48

Entropy = -(p(A) log2(p(A)) + p(B) log2(p(B))) = -(0.4*log2(0.4) + 0.6*log2(0.6))) = 0.97

10. What are the advantages of Random Forests over Decision Tree?
➔ The main advantage of random forest over Decision Trees is that random forest alleviates the problem
of over-fitting. The decision trees are highly likely to overfit the training data if we do not regularize

them by restricting the max depth of the tree, max number of leaf nodes etc. Random forest is an
ensemble technique in which each tree is trained on a bootstrapped sample which is taken from the
training dataset and the prediction is made by taking average of the prediction made by each tree.
So, random forests are very less likely to overfit and they generalize well on unseen data as compared
to decision trees.

11. What is the need of scaling all numerical features in a dataset? Name any two techniques used for scaling.
➔ Most of the times, a dataset will contain features highly varying in magnitudes, units and range. But
since, most of the machine learning algorithms use Euclidean distance between two data points in
their computations, this is a problem because, the results would vary greatly between different units,
5 km and 5000 m. The features with high magnitudes will weight a lot more in the distance calculations
than features with low magnitudes. To avoid this effect, we use scaling which transforms all features
to the same level of magnitudes. The two major scaling techniques used now a days are standardization and normalization

12. Write down some advantages which scaling provides in optimization using gradient descent algorithm.
➔ The advantage which scaling provides in optimization using gradient descent algorithm is that if the
data is scaled, the gradient descent algorithm reached the optimal solution quickly as compared to
the data which is not scaled.

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?
➔ consider the case of a highly imbalanced dataset in which we have 90% data points belonging to label
1 and only 10% belonging to label 0. In that case even if we naively assign label 1 to all the data
points we will get 90% accuracy because 90% data-points are of label 1 so, accuracy here will not be
a good metric to evaluate the performance of a model. Even a very naive model like giving label 1
without even looking at any feature will give you 90% accuracy so, the accuracy will not be a good
metric to use here

14. What is “f-score" metric? Write its mathematical formula.
➔ f-score metric is the metric which to measure the performance of a classification model. F-score (F1
score) is the harmonic mean of precision and recall:

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

15. What is the difference between fit(), transform() and fit_transform()?
➔ fit() : fit() is used for training. It calculates the parameters(weights) to be applied on different features
in order to transform them as required by the model.
transform() : transform uses the parameters(weights) learned by fit() and applies it to respective
features of dataset in order to change them as per requirement of the model.
fit_transform() : fit_transform() first fits() the data and then transforms the same dataset on which fit
is performed. Basically, it does both the tasks on the same dataset.
fit() is generally used on training data in supervised learning where there is no need to change the
training data, fit_transform is used on training data in unsupervised learning where training data must
also be modified and transform() is used only on test data because we need to change the test data
as per the parameters(weights) learned by training.