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? A) GridSearchCV() B) RandomizedCV() C) K-fold Cross Validation D) None of the above

Ans. A

Q2. 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. D

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?

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

Ans. D

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?

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. C

Q5. 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. A

Q6. 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

Q7. 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

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

Ans. B

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.

Ans.

Q10. 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 built 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.

Q11. What is the need of scaling all numerical features in a dataset? Name any two techniques used for scaling

Ans. Feature Scaling or Standardization: It is a step of Data Pre Processing which is applied to independent variables or features of data. It basically helps to normalize the data within a particular range. Sometimes, it also helps in speeding up the calculations in an algorithm.

1) Min Max Scaler

2) Standard Scaler

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

Ans. Gradient descent is an optimization algorithm which is mainly used to find the minimum of a function. In machine learning, gradient descent is used to update parameters in a model. Parameters can vary according to the algorithms, such as coefficients in Linear Regression and weights in Neural Networks.

Let us relate gradient descent with a real-life analogy for better understanding. Think of a valley you would like to descend when you are blind-folded. Any sane human will take a step and look for the slope of the valley, whether it goes up or down. Once you are sure of the downward slope you will follow that and repeat the step again and again until you have descended completely (or reached the minima).

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?

Ans. Applying inappropriate evaluation metrics for model generated using imbalanced data can be dangerous. Imagine our training data is the one illustrated in graph above. If accuracy is used to measure the goodness of a model, a model which classifies all testing samples into “0” will have an excellent accuracy (99.8%), but obviously, this model won’t provide any valuable information for us.

In this case, other alternative evaluation metrics can be applied such as:

* Precision/Specificity: how many selected instances are relevant.
* Recall/Sensitivity: how many relevant instances are selected.
* F1 score: harmonic mean of precision and recall.
* MCC: correlation coefficient between the observed and predicted binary classifications.
* AUC: relation between true-positive rate and false positive rate.

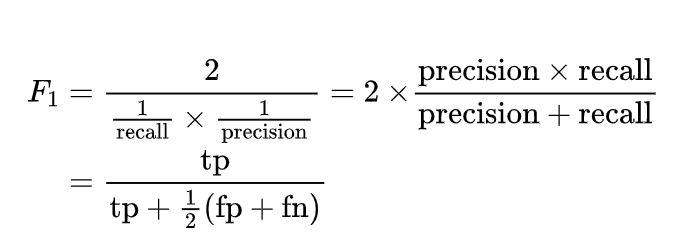
Q14. What is “f-score" metric? Write its mathematical formula.

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](https://deepai.org/machine-learning-glossary-and-terms/classifier) examples into ‘positive’ or ‘negative’.

The F-score is a way of combining the [precision and recall](https://deepai.org/machine-learning-glossary-and-terms/precision-and-recall) of the model, and it is defined as the [harmonic mean](https://deepai.org/machine-learning-glossary-and-terms/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](https://deepai.org/machine-learning-glossary-and-terms/machine-learning) models, in particular in [natural language processing](https://deepai.org/machine-learning-glossary-and-terms/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.



Q15. What is the difference between fit(), transform() and fit\_transform()?

Ans. "fit" computes the mean and std to be used for later scaling. (just a computation), nothing is given to you. "transform" uses a previously computed mean and std to autoscale the data (subtract mean from all values and then divide it by std). "fit\_transform" does both at the same time.