1. Ans-D

2. Ans-D

3. Ans-B

4. Ans-A

5. Ans-C

6. Ans-C

7. Ans-B

8. Ans-C

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

9. Suppose we have a dataset which has 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-Gini index is calculated by subtracting the sum of squared probabilities of each class from one.

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

Ans-Random forest is simply a collection of decision trees whose results are aggregated into one final result.
It reduces overfitting in decision trees and helps to improve the accuracy.It is flexible to both classification and regression problems.It works well with both categorical and continuos values.It automates the missing values in the dataset.,

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

Ans-Scaling can make the difference between a weak machine learning model and a better one.The most common techniques of feature scaling are Normalization and standardization.

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

Ans-Advantages of gradient descent algorithms are its computational efficient.It produces a stable error gradient and a stable convergence.

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-The most common metric used to evaluate the performance of a classification predictive model is classification accuracy.For an imbalanced dataset accuracy is no longer a proper measure,since it does not distinguish between numbers of correctly classified examples of different classes.Henc eit may lead to erroneous conclusion.

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

Ans- F-score is a measure of test’s accuracy.It is calculated from precision and recall of the test.F1 score is the harmonic mean of precision and recall.

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

Ans-fit() calculates the values of parameters.transform function applies the value of the parameters on the actual data and gives the normalized value.The fit transform() function performs both in the same step.same value is got if we perform both steps or in a single step.