Machine Learning 7

1. A, B
2. D
3. B
4. A
5. C
6. C
7. B
8. C

9. Gini index is calculated by subtracting the sum of squared probabilities of each class from one

10. 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 continues values.

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

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

13. 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 doesn't distinguish between numbers of correctly classified examples of different classes. Hence it may lead to erroneous conclusion.

14. 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. 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