8 Multiple Choice questions:

1. Searching in a Binary Tree has a performance of
   1. O(log n)
   2. O(1)
   3. O(n)
   4. O(n/2)
2. Supervised learning’s problems are\_ , and Unsupervised learning’s problems are \_:
   1. Classification and Clustering; Regression & Association
   2. Classification & Association; Regression & Clustering
   3. Classification & Regression; Clustering & Association
   4. Regression & Clustering; Classification & Association
3. Given that there are 15 apples and 20 oranges. Out of the 9 apples that were identified, only 6 are apples and 3 are oranges. The precision and recall rate is:
   1. 6/9 and 6/15
   2. 6/15 and 9/15
   3. 6/9 and 3/9
   4. 3/9 and 6/9
4. The probability of a guessed answer being correct for this question, *P(Correct | not Know)* is 1/4. The probability that you know the correct answer to this question, *P(Know)*, is 2/3. Also, the probability of you answer the question correctly if you know the answer, *P(Correct | Know)*, is 1. If you answer the question correctly, the conditional probability that you know the correct answer, *P (Know | Correct)*, is:

a. 8/9

b. 2/12

c. 3/4

d. 1/3

1. Exploratory Data Analysis is NOT helpful for:
   1. Providing visualization to understand interaction between statistical models
   2. Providing summary statistics for each field in the raw dataset
   3. Identifying interactions between different fields in the dataset
   4. Identifying clusters of similar observations
2. The best way to numerically represent an RGB image is to use:
   1. A vector
   2. A 2D numpy array
   3. A 3D numpy array
   4. A 4D numpy array

6. PCA

7. Word processing: how are topics determined

8. Storytelling with Data / Time Series

True False questions (7)

1. One of the reason for using lists to implement Stack instead of using ‘list’ is:
2. To improve the performance of function Insert and Pop
3. To be able to use more functions that are not available for ‘list’
4. To be able to store different value types like tuple (heterogenous)
5. None of the above (no advantage)
6. Labeled data is required for k-means clustering to work
   1. True
   2. False
7. Overfitting occurs when a model shows low variance and high bias
   1. True
   2. False
8. Edges in an image can be characterized by low value first derivative
   1. True
   2. False
9. Principal component analysis (PCA) can be used with variables of any types: continuous, categorical, or a mixture of these types
   1. True
   2. False
10. With an infinite supply of training data, the trained Naıve Bayes classifier is an optimal classifier
    1. True
    2. False

7.