Worksheet7-Machine Learning

1. A
2. A
3. A
4. C
5. C
6. C
7. A
8. D
9. G(I)=0.45 And H(x)=0.292
10. Random forests consist of multiple single trees each based on a random sample of the training data. They are typically more accurate than single decision trees.
11. The need of scaling is to have all the independent variable to be in the same scale so that it will be easy for algorithms to understand relative relationship better.Two ways of scaling are
    A .Normalization
    B. Standardization
    C.MinMaxScaler

13. Accuracy is not a suitable metric to use when the dataset is imbalanced. F1 is suitable Metric to be used when imbalanced data is there.

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

15. fit() just calculates the parameters and saves them as an internal object's state. transform() method to apply the transformation to any particular set of examples. fit_transform() joins these two steps and is used for the initial fitting of parameters on the training set x, while also returning the transformed x'. Internally, the transformer object just calls first fit() and then transform() on the same data.