Answers for the Statistics

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

10. This is the most common distribution for independent randomly generated variables. This distribution is always normal irrespective of sample size. It's a proper term for the probability bell curve. In this mean is o and standard deviation is 1.

11. One way to handle the missing data is deletion of rows and columns having null values and another way is imputation of values.
    Imputation techniques are
    1. Mean imputation
    2. Substitution
    3. Hot deck imputation
    4. Cold deck imputation
    5. Regression imputation

12. An A/B test is a process whereby hypotheses are made about the relation between two datasets and those datasets are then compared to each other to determine if there is a statistically significant relation or not.

13.No, mean imputation is not a good practice because it ignores the feature correlation.

14. Linear regression analysis used to predict the value of a variable based on the value of another variable. The variable u want to predict is dependent and the variable u r using to predict is an independent variable.

15. Types of statistics
    1. Descriptive
    2. inferential

Machine learning

1. B
2. D
3. D
4. A
5. B
6. D
7. A
8. B
9. D
10. A
11. D
12. A

13. It is calculated by measuring the distance between each data point and its centroid squaring the distance and summing these squares across one cluster.

14. To measure the quality of clustering,we can use silhouette coefficient values of all objects in the dataset.

15. Cluster analysis is a technique that explores naturally occurring groups within a data set known as clusters. It doesn't need to group data points into any predefined groups, which means it is an unsupervised learning method.
    Types
    1. Centroid based
    2. Density based
    3. Distribution based
    4. Hierarchical
    5. Constrant based
    6. Fuzzy clustering