# MACHINE-LEARNING-WORKSHEET-1

1. C
2. D
3. D
4. A
5. D
6. D
7. A
8. B
9. A
10. A
11. D
12. A
13. Cluster Analysis can be calculated with either Mathematical computation or Statistical computation wherein, mathematical computation can be performed by sums of squares using elbow and silhouette method while statistical computation can be performed by evidence against null hypothesis using gap statistic.
14. Clustering quality can be measured by comparing the distinctly grouped data by the model with known data and check whether it falls into correct category or not. We can also check whether model is preventing small clusters which is a sign of good performing model.
15. Clustering is process of grouping similar data points which have relevant characteristics. There are 4 types of Cluster Analysis Techniques:
16. Centroid Clustering
17. Density Clustering
18. Distribution Clustering
19. Connectivity Clustering

# MACHINE-LEARNING-WORKSHEET-2

1. A
2. C
3. A
4. C
5. B
6. B
7. A
8. D
9. A
10. A
11. F
12. E
13. Yes
14. K means uses minimal resources and thereby it is easy to implement and execute.
15. No

# SQL WORKSHEET 1

1. A,D
2. A,B
3. B
4. B
5. A
6. C
7. B
8. B
9. D
10. A
11. A Data warehouse is typically used to connect and analyze business data from heterogeneous sources.
12. The basic difference between OLTP and OLAP is that OLTP is an online database modifying system, whereas, OLAP is an online database query answering system.
13. The major characteristics of Data warehousing are Subject-oriented, Integrated, Time-Variant, Non-Volatile and the functions of Data warehousing are Data consolidation, Data Cleaning, Data Integration.
14. Star Schema comprises of one fact table which would index any number of dimensional tables having fact table at its center and the dimension tables at its peripheral representing the star’s points.
15. Don’t know

# SQL WORKSHEET 2

1. D
2. D
3. A
4. A
5. D
6. B
7. C
8. C
9. B
10. D
11. C
12. C
13. A
14. B,C
15. C,D

# Worksheet\_Statistics\_1

1. A
2. A
3. C
4. D
5. C
6. B
7. A
8. B
9. C
10. The distribution of data which is symmetric about mean and has a bell type curve when represented on x-y axis can be referred as normal distribution.
11. Missing data can be handled in multiple ways like deleting rows in case where the missing values are less and imputing the data when the missing values are moderate.

Different types of imputing techniques are:

* 1. Mean/Median values imputation
  2. Most Frequent value imputation
  3. Imputation using KNN

1. Don’t know

# Worksheet\_Statistics\_2

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