# Machine Learning-2

**1.** a (2 only-Clustering)

**2.** d (1-Regression, 2-Classification, and 4-Reinforcement)

**3.** a (True)

**4.** a (1-Capping & flouring of variables)

**5.** b (1)

**6.** b (No)

**7.** a (Yes)

**8.** d (All of the above)

**9.** a (K-means clustering algorithm)

**10.** d (All of the above)

**11.** d (All of the above)

**12.** Yes, K-means is sensitive to outliers, because a mean is easily influenced by extreme values.

**13.** K-means is better because its easy for implementation. It has few more advantages i.e.,
a. It generalizes to clusters of different shapes and sizes, ex:- elliptical clusters.
b. It scales to large data sets.
c. It can warm start the positions of centroids.
d. Easly adapts to new exam.
e. Guarantees convergence.

**14.** No, it has non-deterministic nature, due to its random selection of data points as initial centroids. This means that if we run the algorithm several times on the same data, we can get different result.

# SQL_Worksheet_2

**1.** D (Unique)

**2.** C (Null)

**3.** A (Each entry in the primary key uniquely identifies each entry or row in the table)

**4.** A (There should not be any duplicate entries)

**5.** B (Foreign key)

**6.** D (1)

**7.** A (one to many)

**8.** C (one to one)

**9.** B (Supplier id)

**10.** C (3)

**11.** B (many to one)

**12.** C (Tables)

**13.** A (Insert into)

**14.** B (Unique) & C (Primary key)

**15.** A (Any blood group can contain one of the following values- A,B,AB,& O) & B (A blood group can only contain characters.)

# Statistics Worksheet-2

**1.** C (Both i.e., SD & Mean)
**2.** C (12)
**3.** D (All of the above)
**4.** C (Both of these i.e., exhaustive & mutually exclusive)
**5.** B (Summarizing & explaining a specific set of data)
**6.** B (Data set)
**7.** A (2 or more)
**8.** B (Scatterplot)
**9.** D (Analysis of Variance)
**10.** A (Z-score)
**11.** C (Mean)
**12.** D (40005.2)
**13.** D (Mean)
**14.** A (Descriptive and Inferences)
**15.** D (H-L)