WORKSHEETS_SET3-MACHINE LEARNING :

1)d

2)d

3)a

4)b

5)d

6)c

7)c

8)

9)

Subjective Questions :

13) Clustering helps in understanding the natural grouping in a dataset. Their purpose is to make sense to partition the data into some group of logical groupings. Clustering quality depends on the methods and the identification of hidden patterns.

14) Clustering performance can easily be improved by applying ICA blind source separation during the graph Laplacian embedding step. Applying unsupervised feature learning to input data using either RICA or SFT, improves clustering performance.

SQL WORKSHEET 3:

```
1) create table if not exists customers(
    customerNumber int not null auto_increment primary key,
    customerName varchar(255),
    contactLastName varchar(255),
    contactFirstName varchar(255),
    phone varchar(15),
    addressLine1 varchar(255),
    addressLine2 varchar(255),
    city varchar(255),
    state varchar(255),
    postalCode varchar(255),
    country varchar(100),
    employeeNumber int not null,
    creditLimit decimal(15, 2),
    foreign key fk_employees(employeeNumber)
    references employees(employeeNumber)
 )
```

```
2) create table if not exists orders(
    orderNumber int auto_increment not null primary key,
    orderDate date,
    requiredDate date,
    shippedDate date,
    statuses text,
    comments text,

    customerNumber int not null,

    foreign key fk_customers(customerNumber)

    references customers(customerNumber)

 )
```

3) SELECT * FROM Orders;

4) SELECT comments FROM Orders;

5)

6)SELECT employeeName,lastName,firstName from employees

STATISTICS WORKSHEET 3 :

1)b

2)c

3)a

4)a

5)c

6)b

7)b

8)d

9)a

10)Bayer's theorem is a theorem describing how the conditional probability of each of a set of possible causes for a given observed outcome can be computed from knowledge of the probability of each cause and the conditional probability of the outcome of each cause.

11) Z-score indicates how much a given value differs from the standard deviation. The Z-score, or standard score, is the number of standard deviations a given data point lies above or below mean.

12) A t-test is a statistical test that compares the means of two samples. It is used in hypothesis testing, with a null hypothesis that the difference in group means is zero and an alternate hypothesis that the difference in group means is different from zero.

13) A percentile is a measure used in statistics indicating the value below which a given percentage of observations in a group of observations fall. For example, the 20th percentile is the value (or score) below which 20% of the observations may be found.

14) Analysis of variance (ANOVA) is an analysis tool used in statistics that splits an observed aggregate variability found inside a data set into two parts: systematic factors and random factors. The systematic factors have a statistical influence on the given data set, while the random factors do not. Analysts use the ANOVA test to determine the influence that independent variables have on the dependent variable in a regression study.

15) ANOVA is helpful for testing three or more variables. It is similar to multiple two-sample t-tests. However, it results in fewer type I errors and is appropriate for a range of issues. ANOVA groups differences by comparing the means of each group and includes spreading out the variance into diverse sources.