MACHINE LEARNING ANSWERS

1 - d. All of the above

2 - d. None

3 - c. Reinforcement learning and Unsupervised learning

4 - b. The tree representing how close the data points are to each other

5 - d. None

6 - c. k-nearest neighbour is same as k-means

7 - d. 1, 2 and 3

8 - a. 1 only

9 - a. 2

10 - c. Predicting whether stock price of a company will increase tomorrow.

11 – c

12 – a

13 - Clustering is useful for exploring data. If there are many cases and no obvious groupings, clustering algorithms can be used to find natural groupings. Clustering can also serve as a useful data-preprocessing step to identify homogeneous groups on which to build supervised models

14 - K-means clustering algorithm can be significantly improved by using a better initialization technique, and by repeating (re-starting) the algorithm. When the data has overlapping clusters, k-means can improve the results of the initialization technique

SQL ANSWERS

1 - CREATE TABLE customers

(customernumber int NOT NULL,

customername char(50) NOT NULL,

contactLastname char(50),

contactFirstname char(50),

address char(50),

phone int,

city char(50),

state char(25),

zip_code char(10)

);

2 - CREATE TABLE order

(ordernumber int NOT NULL,

Orderdate int NOT NULL,

Requireddate int,

Shippeddate int,

status char(50),

comments(50),

```
Foreignkey (customernumber) REFERENCES customer(customernumber));
```

3 – select * from orders;

4 – select comments from orders;

5 – select orderDate,sum(ordernumber) from orders;

6 – select employeNumber, lastName, firstName from employees;

7 – select orderNumber, customerName from customers, orders;

8 – select customername from customers;

9 – select paymentdate,sum(amount) from payment;

10 – select productName, MSRP, productDescription from products;

```
11 - SELECT p.`product_id`, p.`name`,
SUM(o.`quantity`) AS quantity
FROM `Order_Detail` AS o
    INNER JOIN `Product` AS p
    ON o.`product_id` = p.`product_id`
GROUP BY o.`product_id`
ORDER BY SUM(o.`quantity`) DESC, p.`name` ASC
LIMIT 3

12 - SELECT customerid, COUNT(customerid) AS
CountOfCust
FROM orders
GROUP BY cityname

13 - SELECT customerid, COUNT(customerid) AS
CountOfCust
FROM orders
GROUP BY state
14 – select employeenumber,extension from employees;
```

STATISTICS ANSWERS

1 - b) Total Variation = Residual Variation + Regression Variation

2 - c) binomial

3 - a) 2

4 - a) Type-I error

5 - b) Size of the test

6 - a) Decrease

7 - b) Hypothesis

8 - d) All of the mentioned

9 - a) 0

10 - Bayes' theorem provides a way to revise existing predictions or theories (update probabilities) given new or additional evidence

11 - The value of the z-score tells you how many standard deviations you are away from the mean. If a z-score is equal to 0, it is on the 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 comparison score between a particular score and the scores of the rest of a group. It shows the percentage of scores that a particular score surpassed

14 - Analysis of variance (ANOVA) is a statistical technique that is used to check if the means of two or more groups are significantly different from each other. ANOVA checks the impact of one or more factors by comparing the means of different samples

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