# MACHINE-LEARNING-WORKSHEET-3

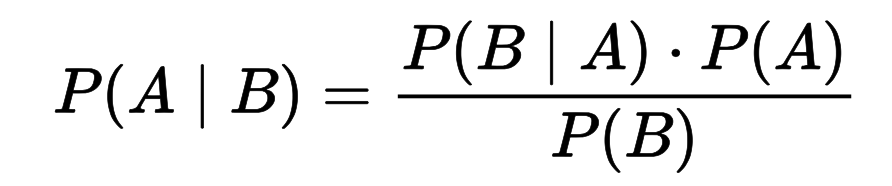
1. D
2. C
3. D
4. B
5. C
6. C
7. D
8. A
9. A
10. B
11. B
12. C
13. If there are no certain groupings in the dataset then clustering can help in separating chunk of data which are similar in characteristics.
14. Profile clustering is generating descriptions of cluster with reference to the input variables used for cluster analysis.
15. A significant parameter which is directly proportional to the performance of clustering is the initialization process and restarting the algorithm. Number of clusters can be changed as a parameter and the process can be reinitialized to get better performance.

# SQL WORKSHEET 3

1. CREATE TABLE CUSTOMERS (CUSTOMERNUMBER INT,CUSTOMERNAME VARCHAR(255),CONTACTLASTNAME VARCHAR(255), CONTACTFIRSTNAME VARCHAR(255), PHONE INT, ADDRESSLINE1 VARCHAR(255), ADDRESSLINE1 VARCHAR(255), CITY VARCHAR(255), STATE VARCHAR(255), POSTALCODE INT, COUNTRY VARCHAR(255), SALESREPEMPLOYEENUMBER INT, CREDITLIMIT INT);
2. CREATE TABLE ORDERS (ORDERNUMBER INT, ORDERDATE DATE, REQUIREDDATE DATE, SHIPPEDDATE DATE, STATUS VARCHAR(255), COMMENTS VARCHAR(255), CUSTOMERNUMBER INT);
3. SELECT \* FROM ORDERS;
4. SELECT COMMENTS FROM ORDERS;
5. SELECT ORDERDATE, COUNT(ORDERNUMBER) FROM ORDERS GROUP BY DATE;
6. SELECT EMPLOYEENUMBER, LASTNAME, FIRSTNAME, FROM EMPLOYEES;
7. SELECT C.CUSTOMERNAME, O.ORDERNUMBER, FROM ORDERS O, CUSTOMERS C WHERE C.CUSTOMERNUMBER=O.CUSTOMERNUMBER;
8. SELECT C.CUSTOMERNAME, E.FIRSTNAME FROM CUSTOMERS C, EMPLOYEES E WHERE C.SALESREPEMPLOYEENUMBER=E.EMPLOYEENUMBER;
9. SELECT PAYMENTDATE, SUM(AMOUNT) FROM PAYMENTS GROUP BY PAYMENTDATE;
10. SELECT PRODUCTNAME, MSRP, PRODUCTDESCRIPTION FROM PRODUCTS;
11. SELECT P.PRODUCTNAME, P.PRODUCTDESCRIPTION, SUM(O.QUANTITYORDERED) TOTAL FROM PRODUCTS P, ORDERDETAILS O WHERE P.PRODUCTCODE=O.PRODUCTCODE GROUP BY PRODUCTCODE ORDER BY TOTAL DESC LIMIT 1;
12. Didn’t get Join for order with city.
13. SELECT STATE, COUNT(CUSTOMERNAME) TOTAL FROM CUSTOMERS GROUP BY STATE ORDER BY TOTAL DESC LIMIT 1;
14. SELECT EMPLOYEENUMBER, FIRSTNAME||’ ‘||LASTNAME FROM EMPLOYEES;
15. SELECT C.CUSTOMERNAME, O.ORDERNUMBER, O1.QUANTITYORDERED\*O1.PRICEOFEACH FROM CUSTOMERS C, ORDERS O, ORDERDETAILS O1 WHERE C.CUSTOMERNUMBER=O.ORDERNUMBER AND O.ORDERNUMBER=O1.ORDERNUMBER;

# STATISTICS WORKSHEET 3

1. B
2. C
3. A
4. A
5. C
6. A
7. B
8. D
9. A
10. Bayes’ theorem is formula to calculate conditional probability



1. Z score is statistical measurement to score relationship with mean of the group of scores.
2. T test is also similar to Z score and it also determines the relationship with mean of datapoints but T test is used when the sample of data is less than 30.
3. A percentile score is comparison between a particular score and the scores of rest of the group. If someone secures 70 percentile marks in a class then that person is ahead of 70 percent students and behind 30 percent students in scoring marks.
4. ANOVA is procedure for testing the difference among different groups for homogeneity.
5. ANOVA can figure out if we can reject null hypothesis. Basically it is a test of checking the statistical significance within the group and with the groups of data.