Objective type

Machine learning assignment-3	Statistics worksheet-3
Ans1) D	Ans1) B
Ans2) D	Ans2) C
Ans3) C	Ans3) A
Ans4) B	Ans4) A
Ans5) D	Ans5) B
Ans6) C	Ans6) B
Ans7) D	Ans7) B
Ans8) A	Ans8) D
Ans9) A	Ans9) A
Ans10) B	
Ans11) A	
Ans12) B	

Subjective type

Machine learning-3

Ans13) the importance of clustering is:

It removes the inefficiency and also helps to starting the local search procedure again. It helps in determining the initial structure of the data. It is used to analyze the model and the region(vector) of attraction. It helps us to understand the natural grouping of the data as it is for making sense in partition of the data into some groups of logical groupings. The quality of this technique however depends on the methods that are we going to use and the identification of hidden patterns. It is also used to detect the outliers present in the dataset.

Ans14) I can improve the cluster performance by making the use of ICA (Independent Component Analysis) blind source separation technique in the cases of graph based. Performing the unsupervised feature learning to input data using RICA OR SFT. The performance can also be achieved by using applying k-means clustering technique on the ICA components when the reduction of PCA dimension on the input data is done but the amount of PCA and ICA components should be limited as to the number of unique classes.

Statistics worksheet-3

Ans10) Bayes' theorem says that the probability that an event A occurs given that another event B has already occurred is equal to the probability that the event B occurs given that A has already occurred multiplied by the probability of occurrence of event A and divided by the probability of occurrence of event B.

Ans11) Z-score is a numerical measurement that describes the relationship of value to the mean of the group of values. Z-score is measured in terms of standard deviations from the mean. If Z-score is 0, then the score of the data points is identical to the score of mean.

Ans12) A t-test is a statistical test that is used to compare the means of two groups. It is often used in hypothesis testing to determine whether a process actually has an effect on the population of interest or the two groups are different from one another.

Ans13) 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. For example, if you score 75 points on a test, and are ranked in the 85th percentile, it means that the score 75 is higher than 85% of the scores. In the other words, a value on a scale of 100 that indicates the percent of a distribution that is equal to or below it a score in the 95th percentile.

The percentile rank is calculated using the formula

$R=P*(N)/100$

Ans14) Analysis of variance or ANOVA is a statistical method that separates observed variance data into different components to use for additional tests. A one-way ANOVA is used for three or more groups of data, to gain information about the relationship between the dependent and independent variables. The formula is:

$F=MSE/MST$

F=ANOVA coefficient, MST=Mean sum of squares due to treatment, MSE=Mean sum of squares due to error.

Ans15) ANOVA test is a way to find out if results obtained from the experiment are significant. In other words, they help you to figure out if you need to reject the null hypothesis or accept the alternate hypothesis. Basically, you're testing groups to see if there's a difference between them. Basically, the ANOVA can help you know whether there are significant differences between the means of your independent variables or not. When you understand how the mean of each independent variable is different from the other one, you can understand which of them has a connection to your dependent variable and can learn about their behavior with each other.

SQL worksheet-3

Ans1) cur.execute('CREATE table customers(customerNumber int primary key, customerName text, contactLastName text, contactFirstName text, phone int, addressLine1 int, addressLine2 text, city text, state text, postalCode int, country text, salesRepEmployeeNumber int, creditLimit int)')

Ans2) cur.execute('CREATE table orders(orderNumber int primary key, orderDate int, requiredDate int, shippedDate int, status text, comments text, customerNumber int, foreign key(customerNumber) references customers(customerNumber)')

Ans3) results=cur.execute('SELECT * FROM orders')

 results.fetchall()

Ans4) results=cur.execute('SELECT comments FROM orders')

 results.fetchall()

Ans5) results=cur.execute('SELECT orderDate, orederNumber FROM orders')

 results.fetchall()

Ans6) results=cur.execute('SELECT employeeNumber, lastName, firstName FROM employees')

 results.fetchall()

Ans7) results=cur.execute('SELECT orders.orderNumber, customers.customerName FROM orders LEFT JOIN customers on orders.customerNumber=customers.customerNumber')

results.fetchall()

Ans8) results=cur.execute('SELECT customerName, salesRepEmployeeNumber FROM customers')

results.fetchall()

Ans9) results=cur.execute('SELECT paymentDate, amount FROM payments')

results.fetchall()

Ans10) results=cur.execute('SELECT productName, MSRP, productDescription FROM products')

results.fetchall()

Ans11) results=cur.execute('SELECT productName, productDescription FROM products INNER JOIN orderdetails ON products.productCode=Orderdetails.productCode GROUP BY products.productCode ORDER BY SUM(orederdetails.quantityOrdered) DESC LIMIT 1;')

results.fetchall()

Ans12) results=cur.execute('SELECT city FROM customers INNER JOIN orders ON customers.customerNumber=orders.customerNumber GROUP BY customers.city ORDER BY SUM(orders.orderNumber) DESC LIMIT 1;')

results.fetchall()

Ans13) results=cur.execute('SELECT state FROM customers GROUP BY customers.state ORDER BY SUM(customers.customerNumber) DESC LIMIT 1;')

results.fetchall()

Ans14) results=cur.execute('SELECT employeeNumber, firstname+' '+last name' AS fullName FROM employees GROUP BY employees.fullName')

results.fetchall()

Ans15) results=cur.execute('SELECT customerName, orderNumber, amount FROM orders INNER JOIN customers USING (customerNumber) INNER JOIN payments USING (customerNumber) ORDER BY payments.amount DESC LIMIT 1;')

results.fetchall()