Set	**:**	**B**
Unique Paper Code	**:**	**32341403**
Course	**:**	**B. Sc. (H) Computer Science**
Paper	**:**	**Database Management Systems**
Semester	**:**	**IV**
Duration	**:**	**2 Hours**
Maximum Marks	**:**	**75**

Attempt any FOUR questions

All questions carry equal marks

(For courses effective from Academic Year 2015-16)

Q1 Consider the following case study:

BookMyEvent is an online booking website which allows customers to book tickets for events, shows and artists' performances.

The site stores details about various performance artists that the customers might be interested in such as the artist's name, gender, date of birth, latest work. Artists are associated to an artist category, namely, comedy, tragedy, theatre, T.V. artist. Each category has a code, a title and a description and most categories have several artists, although some are not populated yet.

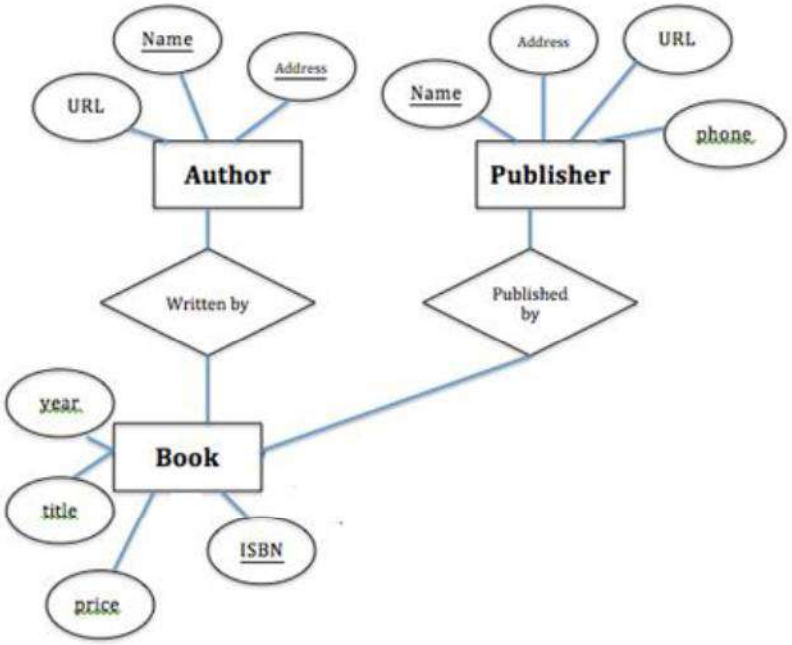
Artists may perform at many different events. All events have an id, a name, a description, start time, and duration. An event may have many news stories released about it, namely, 'new show in city' or 'successfully running for many weeks', to develop interest in the events. Event news has an id, title, author and date.

Events are held at a specific venue. Venues have a name, address, lead contact, seating capacity, star rating. Star ratings are stored with a number, **1-5**.

Many tickets are sold for each event. Tickets include a number, type, such as standard or VIP and a price. Each ticket is for one event only. Customers have to register on the system before they can reserve tickets. They must add standard information including name, address, email, and contact number. A customer can buy many tickets for an event though a ticket is for one event and one customer.

For the above given case study, identify the entities and relationships with their associated attributes, cardinality ratio, and participation constraints. Also, identify subclasses for entities, if they exist. Then, based on these, draw an Enhanced Entity-Relationship Diagram. Clearly state any assumptions you make.

Q2



Study the above ER diagram. For each of the relationships, find the cardinality ratio, and participation constraint values. Map the ER diagram to the relation schema. Clearly show all the steps. Mention all the primary and foreign keys for each relation. Clearly state any assumptions that are made.

Q3 You are hired as a database designer for an **animal adoption agency**. You designed the following relations.

Animal (AnimalID, Name, DateAdmitted, PreviousOwnerID)
Adopter (ID, Name, Address, OtherAnimals)
Adoption (AnimalID, AdoptDate, PrevOwnerID, CurrentAdopterID)
MedicalRecord (AnimalID, age, vaccine, prescription, dosage)

- The underlined attributes are the primary keys.
- The relation **Animal** stores the information about the animal. **DateAdmitted** is of **Date** data type. The **PreviousOwnerID** value should be **NULL** by default. It references the **ID** attribute in the relation **Adopter**.
- The **Adopter** relation stores information about the person who adopts the animal. The attribute **OtherAnimals** should allow values as **Yes** or **No** only.
- The **Adoption** relation depicts the relationship between the relations **Animal** and **Adopter**. **AdoptDate** should be of **Date** data type and should not be **NULL**. **PrevOwnerID** and **CurrentAdopterID** refer to the **ID** attribute of the relation **Adopter** and may or may not be the same.
- **AnimalID** attribute in the **MedicalRecord** relation references the **AnimalID** attribute in the **Animal** relation.

For the above relations, answer the following questions using SQL:

- Write **CREATE TABLE** command for each of the tables. You must use data types, **PRIMARY KEY** constraint, and **FOREIGN KEY** constraint, as applicable. Implement the other constraints mentioned above
- For each of the given relations, write a command to **INSERT** one row with appropriate values

Q4 Consider the relations in question **3**.

Design **four** queries in English language. The **first** query should involve a join and should display the results in a sorted fashion. The **second** query should involve an aggregation operator and the **group by** clause. The **third** query should update the data (based on a condition) in one of the tables. The **fourth** query should delete all the data in one of the tables.

Write the corresponding SQL queries for **any of the three** English queries.
Write relational algebra queries for **any one of the** English queries.

Example: If you were to write a query using **WHERE** clause, you might answer as follows:
English query: ***Find name of the animal whose AnimalID is 234567.***
SQL query:

```
SELECT Name
FROM Animal
WHERE AnimalID = 234567;
```

Q5 Consider a relation **R (A, B, C, D, E, F)** with the following set of functional dependencies

$$F = \{AB \rightarrow C,\ DC \rightarrow AE,\ E \rightarrow F\}$$

Find any two keys for the given relation? Is this relation in **2NF**? If not, state the reason and decompose this relation until each of the decomposed relations are in **2NF**.

Q6 Consider a disk with block size **B = 1024 bytes**. A block pointer is **P = 8 bytes** long and a record pointer is **P_R = 10 bytes** long. A file has **r = 102,000 Book** records of fixed length. Each record has the following fields: **BookID (6 bytes), Name (30 bytes), Author (30 bytes), Publisher (30 bytes), Year (2 bytes), PrintedCopies (2 bytes).** Assume that the file is ordered by the key field **BookID**.

- Calculate the record size **R** for the file in bytes.
- Assuming the un-spanned organization, calculate the blocking factor and the number of file blocks **b** required to store all the records.
- Calculate the number of block accesses required for searching a record in the data file using binary search.

Suppose that we construct a primary index on **BookID** for the above file.

- Calculate the size of an entry in the index R_i.
- Calculate the index blocking factor bfr_i
- Find the total number of index entries r_i and the number of index blocks, b_i.
- Calculate the number of block accesses required for searching a record in the data file using this primary index.

Unique Paper Code	**:**	**32341401**
Name of the Course	**:**	**B. Sc. (Hons.) Computer Science - CBCS**
Name of the Paper	**:**	**Design and Analysis of Algorithms**
Semester	**:**	**IV**
Duration	**:**	**2 Hours**
Maximum Marks	**:**	**75**
Year of Admission	**:**	**2015, 2016 & 2017**

Instructions for Candidates:
Attempt Any Four questions. All Questions carry equal marks.

Q1. Following table is a record of minimum and maximum temperatures of 10 cities:

S.No.	City Name	Min Temp (in °C)	Max Temp(in °C)
1	Delhi	26	39
2	Mumbai	28	34
3	Srinagar	02	8
4	Chennai	29	38
5	Bangalore	22	33
6	Kolkata	27	35
7	Shillong	16	21
8	Hyderabad	25	40
9	Patna	24	36
10	Shimla	09	15

Using the most efficient algorithm sort the cities in the increasing order of minimum temperature. Identify the algorithm that can be used to find the city with the 5th maximum temperature without sorting this data again. Apply this algorithm on the above data to find that city, exhibiting the worst-case performance of algorithm.

Q2. Build a Red Black Tree using the following keys

15,30,45,60,75,90

If x denotes the value of the root of tree built with above values then delete x from this tree. Insert the value x again in this tree. Is this tree same as the one before deletion of x? Justify.

Q3. Suppose divide and conquer techniques is used to solve a problem on an array A of size n. The algorithm divides the problem into 'a' subproblems each of size n/b. It solves the problem recursively on these subproblems and combine their solution to construct the solution of the original problem. Let P(n) denotes the cost of dividing the problem into subproblem and M(n) denotes the cost of combining the solutions then write a recursive function to find running time of the algorithm. Give P(n) and M(n) for both Quicksort

and Mergesort. Derive the recursive function to find running time of the following algorithm. Solve this function and derive complexity in terms of asymptotic notation.

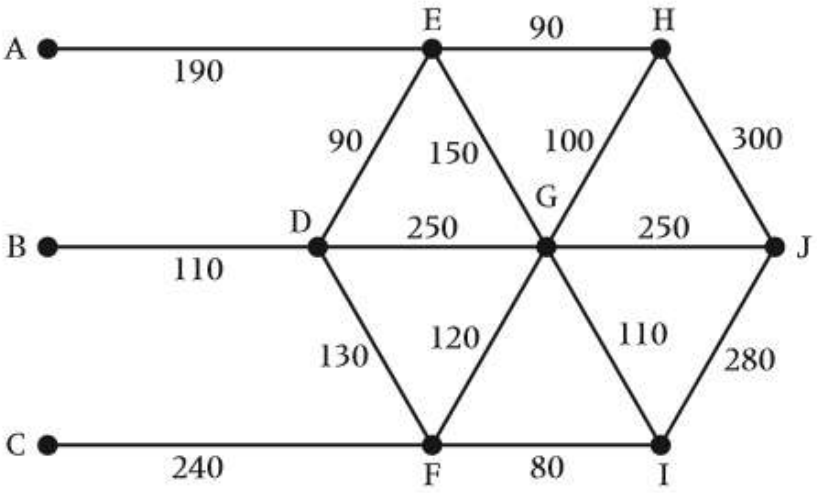
```
My_function(x, a, c, b)
{
  if(x==1)
  {    Print a,c;
       return ();
  }
  else
  {
          My_function(x-1, a, b, c) ;
          print a,c;
          My_function (x-1, b, c, a);
  }
}
```

Q4. A farmer needs to sell his products in the market which is across the river. He has packed each product in separate jute bags and labelled each jute bag a unique number along with its weight and price. He hires a boat to carry these bags to market. But boatman puts a condition. His boat can carry weight of 110 Kg only. Boatman and farmer together weigh 102 Kg. Now the farmer is in a fix - he needs to select the jute bags in such a way that he gets the maximum price in the market without violating the capacity of boat. Write an algorithm to find the optimal solution for farmer's problem. Using that algorithm, find the solution for the following instance, illustrating each step clearly.

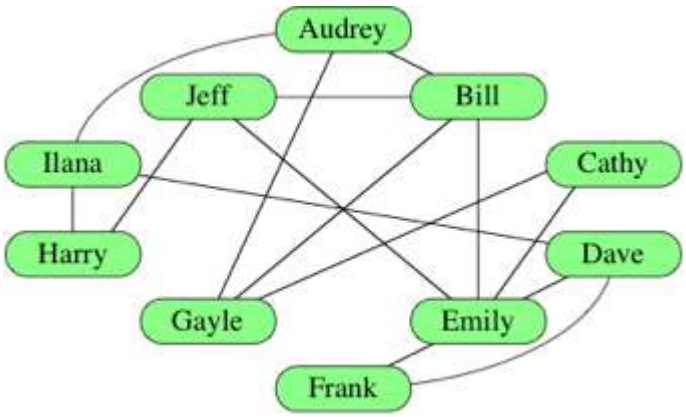
Jute Bag Number	Product in bag	Weight of jute bag (Kg)	Price (Rs)
1	Onions	2	63
2	Potatoes	2	80
3	Maize	3	95
4	Rice	4	430
5	Wheat	5	450

Q5. The following diagram shows the network of roads that connect the head office of Jayshree Bank (indicated by J) to its three branches A, B and C. The number on each edge represents the distance in kilometres, along each road. Daily in the morning cash is sent to the branches from head office in a single delivery van. After delivery, the van does not return to head office. It remains stationed at the branch in which last delivery was made. In the evening delivery van collects the cash from the branch where it was stationed and then goes to other two branches and collects the cash from there and return to head office following the same path. For security reasons, the delivery van must take the shortest path in the morning as well as in the evening. Compute the sequence in which it must visit the branches in the morning starting from head office. Consider that after reaching first

branch it goes directly to second branch and then to third branch where it gets stationed. It does not go back to head office in between. Show the complete shortest path that the van should take starting from J to the branch where it will be stationed.



Q6 Following diagram represents a social network. A line between two people means that they are connected on a meeting app. Message broadcasting is a chain reaction in this network. When a person broadcasts a message to all his contacts, his contacts in turn broadcast the message to all their contacts and so on. Write an algorithm to check whether the message broadcasted by any given person 'x' in the network reaches to each and every other person in the network. In the following instance show the path of transmission from Audrey to all other persons using the algorithm. Also check whether it is possible to divide people in this social network in two groups such that people belonging to a group are not connected on the social media.



Unique Paper Code : 32341402
Name of Paper : Software Engineering
Name of Course : B.Sc. (H) Computer Science (CBCS)
Semester : IV

Duration : 2 Hours **Max Marks : 75**

Attempt any four questions.
All Questions carry equal marks.

Q1,Q2 and Q3 are based on the case study given below:

A Company XYZ Ltd. has been assigned the project for development of an e-TRS (electronic-Ticket Railway System) for reservation of railway tickets. The different processes in e-TRS system are **Fill Journey Details, Validate Details, Get Fare, Pay On-line and Allocate Seat.**

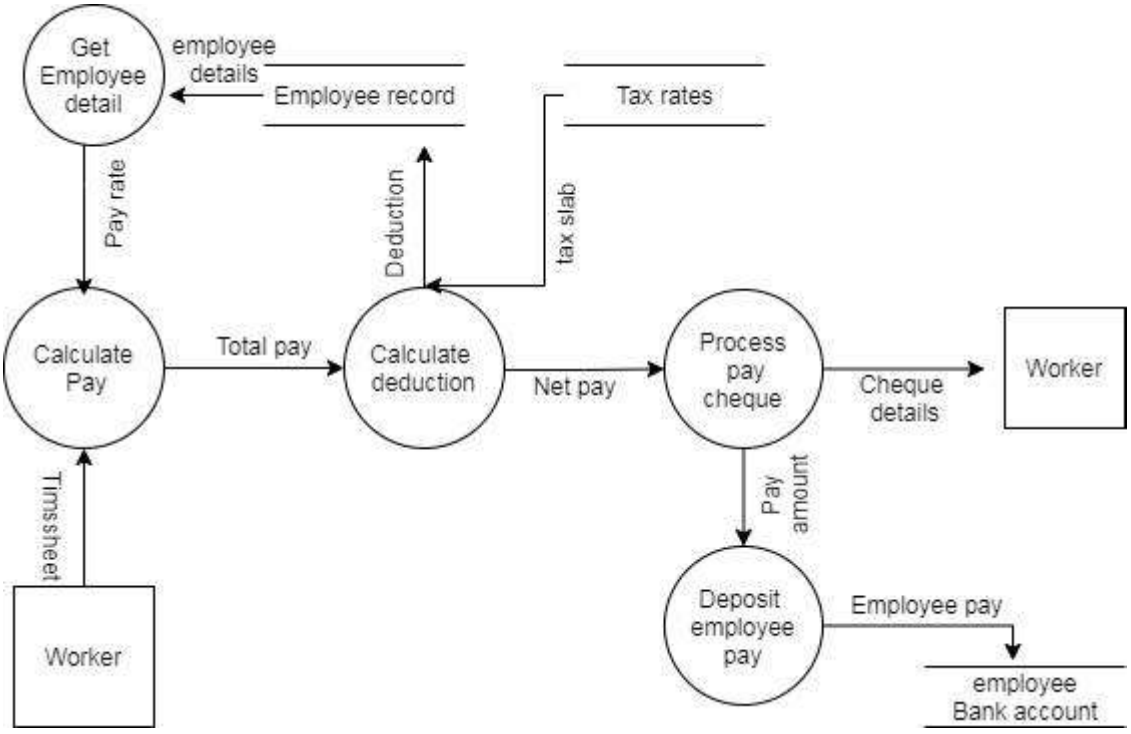
- The internet users register on web site, view seat/berths availability, fill form and book tickets, cancel tickets and other related functions.
- After registration the user can interact with e-TRS System by their user-id and password.
- To book the seats the user is required to fill the journey details such as train number/name, number and name of passenger/s, journey date and names of source and destination stations.
- Before checking the availability of seats and booking them the system should validate the journey details entered by the user.
- In case of non-availability of the tickets the system should suggest the seats availability in the trains for next three days.
- The system should compute total ticket fare for journey details given by user.
- Train administrator should be able to maintain information of fare details, seat/berths class wise of different trains, booked tickets and list of passengers booked.
- The system must be able to accommodate at least 3000 concurrent web users.

Q1. What are the possible software development process models that can be used for e-TRS system described in the case study given above? Select the most suitable model from possible models. Justify your answer.

Q2. Develop a good quality SRS (Software Requirement Specification) Document as per IEEE Standard for the system described in the case study given above. The SRS must include at least two functional, non-functional, performance and User Interfaces requirements each.

Q3. Draw Context level, level 1 and level 2 Data Flow Diagrams for the system described in the case study given above. The diagrams should be clearly labeled with input, processing and data entities.

Q4. Identify the Transform Centre Process using the level 1 Data Flow Diagram for Payroll System given below. Use the transform mapping steps to build the Functional (Call and Return) Architectural Design first level and second level factoring of the DFD.



Q5. Compute the Function Point value for a project with the following parameter values:

Measurement Parameters	Count	Weighing factors		
		Low	Average	High
Number of user inputs	33	3	4	6
Number of user outputs	60	4	5	7
Number of user inquiries	24	3	4	6
Number of files	6	7	10	15
Number of external interfaces	3	5	7	10

Assume that the measurement parameters are equally divided among low, average and high complexity and the complexity adjustment value is 1.05.

Explain the use of DRE as a quality metric for quality assurance and control activities? Compute DRE for the modeling activity given that 30 errors were introduced during the modeling activity and 12 errors were uncovered during the construction activities that were traceable to errors that were not discovered in the modeling activity.

Q6. Draw a flow graph of the program given below that computes the exponentiation power of a variable. Mark the regions in the flow graph and identify the number of independent paths in the flow graph. Also determine the Cyclomatic Complexity of the flow graph using three different methods.

```
1.    int a, b, exp; float c;
2.    scanf("%d %d",&a,&b);
3.    if(a<0)
4.       exp=-a;
5.    else
6.       exp=a;
7.    c=1.0;
8.    while(exp!=0)
9.       {
10        c=c*b;
11.       exp=exp-1;
12.      }
13.   if(a<0)
14.      c=1.0/c;
15.   printf("%f", c);
```

Unique Paper Code	: 32341602
Name of the Course	: BSc (H) Computer Science
Name of the Paper	: Computer Graphics
Semester	: VI
Year of Admission	: 2015, 2016, 2017
Duration	: 2 Hours
Maximum Marks	: 75

INSTRUCTION:

Attempt any **Four** Questions out of **Six**. All Question Carry Equal Marks.

1. Describe the functioning of a plasma display panel which does not require any refreshing, gives very stable image at the output panel and allows erasing and writing selectively at some finite speed about 21.5 microsecond per unit cell.
2. A dashed line joining two points P(x_1, y_1) and Q(x_2, y_2) has to be plotted. Describe an algorithm to draw a dashed line for joining the points P(x_1, y_1) and Q(x_2, y_2). The length of dash is *p* pixels and length of gap between two dash is *q* pixels.
3. Find a transformation matrix of a triangle having co-ordinates A(2, 0), B(0, 2), C(1, 1) which is rotated an angle **θ** in anticlockwise direction about the origin. Also give the transformation matrix when rotation by an angle **θ** is done about an arbitrary point P(h, k) and then translating by two units in each of x and y directions.
4. From the standard perspective projection point of view, what is the projected image of
    - the point P(p, q, r) on to the plane z= k_z, assume own center of projection.
    - the line segment joining P(2, -2, -6k_z) to Q(6, 2, 0) in the plane z= k_z.

5. The two curves are given as z(t)=(t^2+2t-2, t^2) and r(t)=(t^2+2t+1, t+1). Show that these two curves are both C^0 and G^0 continuous when they meet at z(1) = r(0). Do these curves satisfy C^1 and G^1 continuity? Justify your answer.
6. Construct a Bezier curve of order **three** using four control points which are the corners of the polygon defined by A(2, 2), B(3, 4), C(5, 4) and D(7, 5). Find the co-ordinates for t = 0.1, 0.4, 0.6, 0. 8.

Unique Paper Code : **32347611**
Name of the Paper : **Data Mining**
Name of the Course : **B.Sc. (H) Computer Science**
Semester : **Semester -VI**
Duration of Examination : **Two Hours**
Maximum Marks : **75 Marks**
Year of Admission : **2015, 2016, 2017**

Instructions for Candidates

Attempt Any Four questions. All Questions carry equal marks.

Q 1. Given the following table, classify all the attributes appearing in the table as *binary, discrete or continuous*. Also classify them as qualitative (*nominal or ordinal*) or quantitative (r*atio or interval*). Justify your answer in each case. Show the normalization (scaling data between 0 and 1) of values in the age attribute column. How can you handle missing values in Age column and Height column? Replace Y and N respectively by 1 and 0 in the first column. Replace M and F respectively by 0 and 1 in the second column. You will get one binay vector each for Smoker attribute and Gender attribute. Find out the similarity measure between these two vectors using Jaccard coefficient.

Smoker	Gender	Age	Height	Marital Status
Y	F	32	Tall	Married
Y	M	34	Medium	Marries
N	F	39	Medium	Single
Y	M	41	Tall	Single
Y	M	25	Tall	Divorcee
N	M	36	Tall	Single
Y	F	45	Short	Married
Y	M	31	Tall	Single
N	M	29	Medium	Divorcee
N	F	51	Tall	Single
Y	F	38	Short	Married

Q 2. Given the following binary classification problem :

Instance	A1	A2	Target Class
1	T	T	+
2	T	T	+
3	T	F	-
4	F	F	+
5	F	T	-
6	F	T	-
7	F	F	-
8	T	F	+
9	F	T	-
10	T	F	-

Calculate separately the information gain when splitting is done on A1 and on A2. Which attribute would the decision tree induction algorithm choose? Calculate separately the gain in the Gini index when splitting is done on A1 and A2. Which attribute would the decision tree induction algorithm choose? Is it possible that information gain and the gain in Gini index favour different attributes? Explain your answer.

Q 3. Consider the one dimensional labeled data set given below:

X:	0.5	3.0	4.5	4.6	4.9	5.2	5.3	5.5	7.0	9.5
Y:	-	-	+	+	-	-	+	+	-	-

Classify the data point x= 5.0 according to its 3- and 5- nearest neighbour using majority vote.

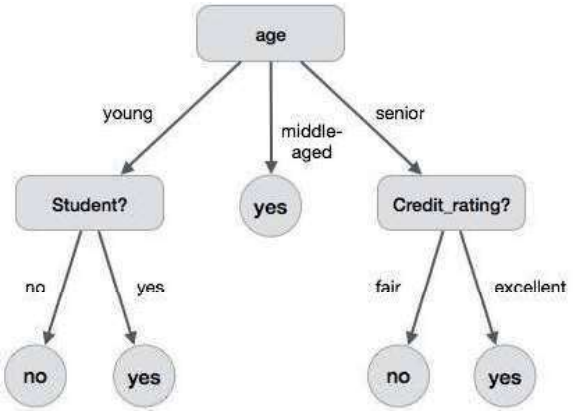
Suppose that there are a total of 60 data mining related documents in a library of 200 documents. Suppose that a search engine retrieves 20 documents after a user enters 'data mining' as a query, of which 5 are data mining related documents. What are the precision and recall?

Q 4. Consider the Market Basket dataset shown below:

Transaction ID	Items Bought
0001	{a,d,e}
0024	{a,b,c,e}
0012	{a,b,d,e}
0031	{a,c,d,e}
0015	{b,c,e}
0022	{b,d,e}
0029	{c,d}
0040	{a,b,c}
0033	{a,d,e}
0038	{a,b,e}

Compute support for item sets {e}, {b,d}, {b,d,e}, {a,b,c,e} and {a,b,c,d,e}. Find all association rules that can be generated from item set {b,d,e} along with the confidence of each rule. Is confidence a symmetric measure?

Q 5. Generate all the rules given the following decision tree:



Arrange the generated rule according to class based ordering and classify the following tuples:

Age = young	student = no	loan_approval=?
Age = senior	credit_rating = excellent	loan_approval=?

Consider a training set that contains 90 positive examples and 300 negative examples. For each of the following rules:

R1: A----->+ (Covers 30 positive and 10 negative examples)

R2:B---->+ (Covers 90 positive and 80 negative examples)

R3:C ---->+ (Covers 4 positive and 1 negative example)

Determine which is the best and which is the worst candidate rule according to rule Accuracy.

Q 6. Use K-means algorithm and Euclidean distance to cluster five data points (A4-A8) given below, into 3 clusters. The coordinates of the data points are:

A1(2,8), A2(2,5), A3(1,2), A4(5,8), A5(7,3), A6(6,4), A7(8,4), A8(4,7).

Use A1, A2, A3 as initial centroids. For which situations K-mean clustering will give good results and when will it fail to produce good results?

Unique Paper Code: 32347607
Name of the Course: B. Sc. (H) Computer Science DSE-3
Name of the Paper: Machine Learning
Semester: VI
Year of Admission: 2015, 2016, 2017

Duration: 2 Hours Maximum Marks: 75

Instructions for Candidates
Attempt any four questions. All questions carry equal marks.

Q1. Mention the assumption of the Naïve Bayes classifier. What are the advantages of the assumption? Consider the following table of ten weather conditions, with the attribute *Play* as the class label.

Outlook	Temperature	Humidity	Wind	Play
Rainy	Hot	High	No	No
Rainy	Hot	High	Yes	No
Overcast	Hot	Normal	No	Yes
Sunny	Mild	Normal	No	Yes
Sunny	Cool	Normal	No	Yes
Sunny	Cool	Normal	Yes	No
Overcast	Cool	High	Yes	Yes
Rainy	Mild	High	No	No
Rainy	Cool	Normal	No	Yes
Sunny	Mild	Normal	No	Yes

Train a Naïve Bayes classifier using the above table and predict the class label *Play* for the given novel instance: Outlook = Rainy, Temperature = Cool, Humidity = High, Wind = Yes. Mention all the prior and the conditional probabilities considered to calculate the posterior probabilities.

Q2. Given the dataset of heights and weights of six individuals, fit a linear regression line using the ordinary least squares method. Find the residual error for the best fit line. Also predict the weight for height 67.

Height (x)	63	66	69	71	71	75
Weight (y)	127	142	162	156	169	208

Q3. Consider the following multivariate linear regression problem. Given the initial values of the regression coefficients $\beta_0, \beta_1, and\ \beta_2$ as 1, 1, and 1, find the cost, J(β). Given the learning rate $\alpha = 0.10$, compute the next set of values for the regression coefficients using the gradient descent method. What is the cost, J(β), for the new values of the coefficients?

X1	X2	Y
0	1	4
1	2	7
2	2	8
3	1	7
2	1	6

Q4. What is regularization? What is the effect of the following on the model?

a) The regularization parameter (λ) is zero
b) The regularization parameter (λ) is very large

With the help of a neat diagram illustrate the following scenarios for a machine learning model:

a) High bias low variance
b) Low bias high variance
c) Low bias low variance

Q5. What is the cost (loss) function for logistic regression? Why is the mean squared error not suitable as a cost function in logistic regression? Certain health risk factors such as high blood pressure and cigarette smoking etc. lead to sudden death. Therefore a multiple logistic regression model was fit with regression coefficients as shown below.

Risk Factor	**Regression Coefficient**
Constant term (β_0)	-15.3
Blood Pressure (mm Hg) (β_1)	0.099
Weight (Kg) (β_2)	-0.0060
Cholesterol (mg/100 mL) (β_3)	0.0056
Glucose (mg/100 mL) (β_4)	0.0066
Smoking (cigarettes/day) (β_5)	0.0069
Age (years) (β_6)	0.0686

Predict the probability of death for a 50 year old man with diastolic blood pressure of 120 mmHg, a relative weight of 100 Kg of study mean, a cholesterol level of 250 mg/100mL, a glucose level of 100 mg/100mL who smokes 10 cigarettes per day. Also, predict the probability of death if diastolic blood pressure is 180 mmHg with other conditions remaining same.

Q6. Design a neural network for the boolean function *AND* for two variables. With the help of a neat diagram show that a single perceptron cannot model non-linear relationships.