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, $J(\beta)$. 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, $J(\beta)$, 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.