

University of Mumbai
Examination June 2021

Examinations Commencing from 1st June 2021

Program: Computer Engineering

Curriculum Scheme: Rev2016

Examination: TE Semester VI

Course Code: CSDLO6021 and Course Name: Machine Learning

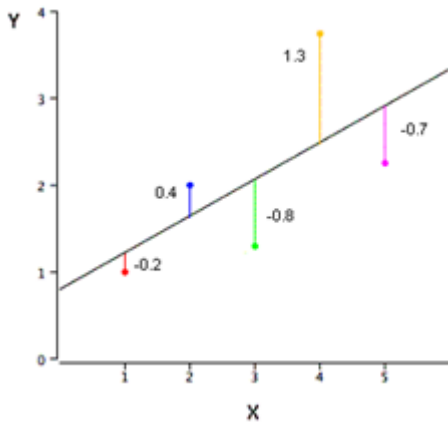
Time: 2 hour

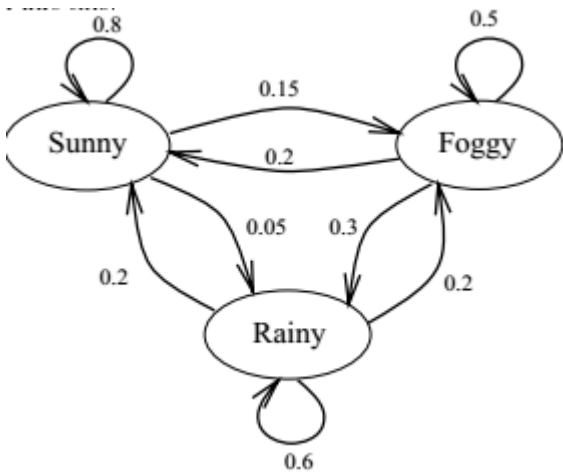
Max. Marks: 80

=====

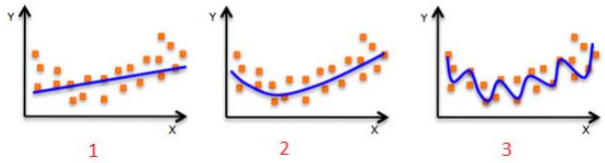
Q1.	Choose the correct option for following questions. All the Questions are compulsory and carry equal marks
1.	Which of the following are examples of unsupervised learning? i. Modeling a spam filter from a set of labeled emails as spam and not spam ii. Given a set of news articles found on the web, group them into articles under different categories iii. Given a database of customer data, automatically discover market segments and group customers into different market segments iv. Given a database of patients diagnosed as either having diabetes or not, learn to classify new patients as having diabetes or not
Option A:	Both i and iv
Option B:	Both i and iii
Option C:	Both ii and iii
Option D:	Both iii and iv
2.	Which of the following options are true about Machine Learning? 1. Machine learning is automatic learning based on experience 2. Machine learning is programmed so that it learns, and past experience is not required. 3. It can learn and improve from the past experience without being explicitly programmed. 4. Machines can learn from past experience, but it must be explicitly programmed.
Option A:	1 and 2
Option B:	2 and 4
Option C:	1 and 4
Option D:	3 and 4
3.	Which of the following is an example of reinforcement learning?
Option A:	Stock price prediction
Option B:	Sentiment analysis
Option C:	Customer segmentation

Option D:	Robot in a maze
4.	In Downhill Simplex method, if $f(x)$ at the reflected point is greater than $f(x)$ at worst point (N) then the new point is obtained by
Option A:	Contraction
Option B:	Multiple Reflection
Option C:	Expansion
Option D:	Multiple contraction
5.	In classical Newton's Method, having Hessian Matrix H, Gradient G, X_{K+1} is computed using
Option A:	$X_{K+1} = X_K + H_K^{-1} * G_K$
Option B:	$X_{K+1} = X_K - H_K^t * G_K$
Option C:	$X_{K+1} = X_K - H_K * G_K$
Option D:	$X_{K+1} = X_K + H_K * G_K$
6.	Which of the following is not true about the derivative free techniques?
Option A:	They use evolutionary concepts.
Option B:	The objective function has to be differentiable
Option C:	These methods use an empirical approach for analysis.
Option D:	Random search and Downhill Simplex are examples of Derivative free techniques.
7.	Given $X=[1 \ 2 \ 3 \ 4]$ $W=[1 \ 1 \ -1 \ -1]$ compute $f(\text{net})$ given $\lambda = 0.5$ using i. Bipolar continuous ii. Unipolar continuous activation function
Option A:	i. 0.7615 ii. 0.880
Option B:	i. 0.880 ii. 0.7615
Option C:	i. -0.7615 ii. 0.1192
Option D:	i. 0.119 ii. -0.7615
8.	Hebbian learning is an example of _____ and perceptron learning is an example of _____
Option A:	Feedforward supervised learning, supervised binary response
Option B:	Feedforward unsupervised learning, supervised binary response
Option C:	Feedback supervised learning, unsupervised binary response
Option D:	Feedback unsupervised learning, supervised multivariate response
9.	_____ is a type of learning rule which works with a layer of neurons.
Option A:	Perceptron
Option B:	Hebbian
Option C:	Widrow Hoff
Option D:	Winner takes all
10.	Which of these statements are false with respect to the metrics in linear regression? a. For a strong linear regression R^2 value should be high b. Multiple R value of 1 represents perfect positive relationship

	c. Karl pearson value of -1 indicates total negative linear correlation d. High value of Sum of Squared Errors(SSE) indicates perfect fit									
Option A:	Both A and B are false									
Option B:	Both A and C are false									
Option C:	Both B and C are false									
Option D:	Only D is false									
11.	The graph below represents a regression line predicting Y from X. The values on the graph shows the residuals for each predicted value. Use this information to compute the Sum of squared errors (SSE) 									
Option A:	4.02									
Option B:	3.02									
Option C:	1.01									
Option D:	0									
12.	<table><tr><td></td><td>Actual True</td><td>Actual False</td></tr><tr><td>Predicted True</td><td>156</td><td>20</td></tr><tr><td>Predicted False</td><td>14</td><td>50</td></tr></table> Compute the specificity and the precision?		Actual True	Actual False	Predicted True	156	20	Predicted False	14	50
	Actual True	Actual False								
Predicted True	156	20								
Predicted False	14	50								
Option A:	Specificity = 88.6% Precision = 71.4%									
Option B:	Specificity = 71.4 % Precision = 88.6%									
Option C:	Specificity = 28.5% Precision = 11.36%									
Option D:	Specificity = 71.4% Precision = 11.36%									
13.	Which is not true statement about Kernel Trick									
Option A:	A Kernel Trick is a method where a Non Linear data is projected onto a higher dimension space so as to make it easier to classify the data where it could be linearly divided by a plane.									

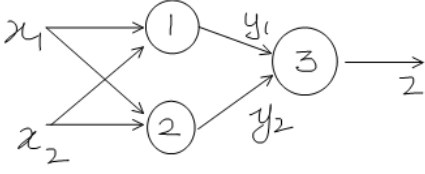
Option B:	A Kernel Trick is a method of transforming the original (non-linear) input data into a higher dimensional space (as a linear representation of data).
Option C:	The Kernel Trick allows us to take linear Support Vector Machines and extend their functionality to classify non-linear data sets.
Option D:	A Kernel Trick is a method which can easily separates the data points in a lower dimensionality space
14.	The difference between naïve Bayesian classifier and Bayesian belief networks is
Option A:	The joint conditional probability distributions are considered in Bayesian Belief networks
Option B:	The joint conditional probability distribution is not considered in Bayesian Belief networks
Option C:	Class conditional independence is always considered in Bayesian Belief networks
Option D:	Class conditional independence is sometimes considered in Bayesian Belief Networks
15.	Today's weather Tomorrow's weather Initial Probability values Sunny 0.25 Rainy 0.75 Foggy 0.30  <pre> graph TD Sunny((Sunny)) -- 0.8 --> Sunny Sunny -- 0.2 --> Rainy((Rainy)) Rainy -- 0.6 --> Rainy Rainy -- 0.05 --> Sunny Rainy -- 0.3 --> Foggy((Foggy)) Foggy -- 0.5 --> Foggy Foggy -- 0.2 --> Sunny Sunny -- 0.15 --> Foggy </pre> Given that today is sunny what is the probability that tomorrow is sunny and the day after is rainy
Option A:	0.01
Option B:	0.004
Option C:	0.04
Option D:	0.32

16.	What is true about Markov Property I. Markov Property is very useful for explaining events, and it cannot be the true model of the underlying situation in most cases. II. The state of the system at time $t+1$ depends only on the state of the system at time t III. The advantages of Markov property are complexity and forecasting accuracy. IV. Markov property is used to forecast the value of a variable whose predicted value is influenced only by its current state
Option A:	i and ii
Option B:	ii and iii
Option C:	ii and iv
Option D:	iii and iv
17.	A square matrix is _____ if all eigen values are _____ Positive definite, Positive Negative definite, Negative Positive definite, Negative Negative definite, positive
Option A:	Both ii and i are correct
Option B:	Both iii and iv are correct
Option C:	All four options are wrong
Option D:	Either iii or iv is right
18.	Identify the correct options regarding Principal Component Analysis (a) Principal component analysis (PCA) can be used with variables of any mathematical types: quantitative, qualitative, or a mixture of these types (b) The major principal component axis has dimensions having the maximum variance. (c) The major principal component axis has dimensions having the minimum variance (d) The most information is retained among the top few principal axes.
Option A:	Both a and b
Option B:	Both b and d
Option C:	Both a and d
Option D:	Both c and d

19.	Compute the eigen values for matrix $A = \begin{bmatrix} 7 & 3 \\ 3 & -1 \end{bmatrix}$
Option A:	$\lambda_1 = 8; \lambda_2 = -2$
Option B:	$\lambda_1 = -8; \lambda_2 = 2$
Option C:	$\lambda_1 = 4; \lambda_2 = -4$
Option D:	$\lambda_1 = -4; \lambda_2 = 4$
20.	 In the graphs 1, 2 and 3 which is best fitted and which is overfitted?
Option A:	2 is best-fitted and 1 is over-fitted
Option B:	1 is best-fitted and 2 is over-fitted
Option C:	2 is best-fitted and 3 is over-fitted
Option D:	1 is best-fitted and 3 is over-fitted

Q2 (20 Marks Each)																																																																															
A	Solve any Two 5 marks each																																																																														
i.	Why is the Support Vector Machine(SVM) called the maximum margin classifier?Explain mathematically the formulation of margin.																																																																														
ii.	What is a saddle point? Minimize $f(x)=x_1^2+x_2^2+2x_1x_2$, with starting initial point X_0 is [0.5,-0.1] (Perform 2 iteration only)using the steepest descent method																																																																														
iii.	What are the steps in designing a Machine Learning Application																																																																														
B	Solve any One 10 marks each																																																																														
i.	<i>Two questions of 10 marks each have to be asked</i> For the following data, to construct the decision tree calculate Gini indexes and determine which attribute is the root attribute. (4) <table><tr><th>Sr. No</th><th>Income</th><th>Defaulting Level</th><th>Credit Score</th><th>Location</th><th>Give Loan?</th></tr><tr><td>1</td><td>low</td><td>high</td><td>high</td><td>bad</td><td>no</td></tr><tr><td>2</td><td>low</td><td>high</td><td>high</td><td>good</td><td>no</td></tr><tr><td>3</td><td>high</td><td>high</td><td>high</td><td>bad</td><td>yes</td></tr><tr><td>4</td><td>medium</td><td>medium</td><td>high</td><td>bad</td><td>yes</td></tr><tr><td>5</td><td>medium</td><td>low</td><td>low</td><td>bad</td><td>no</td></tr><tr><td>6</td><td>medium</td><td>low</td><td>low</td><td>good</td><td>yes</td></tr><tr><td>7</td><td>high</td><td>low</td><td>low</td><td>good</td><td>yes</td></tr><tr><td>8</td><td>low</td><td>medium</td><td>high</td><td>bad</td><td>no</td></tr><tr><td>9</td><td>low</td><td>low</td><td>low</td><td>bad</td><td>no</td></tr><tr><td>10</td><td>medium</td><td>medium</td><td>low</td><td>bad</td><td>no</td></tr><tr><td>11</td><td>low</td><td>medium</td><td>low</td><td>good</td><td>yes</td></tr><tr><td>12</td><td>high</td><td>medium</td><td>high</td><td>good</td><td>yes</td></tr></table>	Sr. No	Income	Defaulting Level	Credit Score	Location	Give Loan?	1	low	high	high	bad	no	2	low	high	high	good	no	3	high	high	high	bad	yes	4	medium	medium	high	bad	yes	5	medium	low	low	bad	no	6	medium	low	low	good	yes	7	high	low	low	good	yes	8	low	medium	high	bad	no	9	low	low	low	bad	no	10	medium	medium	low	bad	no	11	low	medium	low	good	yes	12	high	medium	high	good	yes
Sr. No	Income	Defaulting Level	Credit Score	Location	Give Loan?																																																																										
1	low	high	high	bad	no																																																																										
2	low	high	high	good	no																																																																										
3	high	high	high	bad	yes																																																																										
4	medium	medium	high	bad	yes																																																																										
5	medium	low	low	bad	no																																																																										
6	medium	low	low	good	yes																																																																										
7	high	low	low	good	yes																																																																										
8	low	medium	high	bad	no																																																																										
9	low	low	low	bad	no																																																																										
10	medium	medium	low	bad	no																																																																										
11	low	medium	low	good	yes																																																																										
12	high	medium	high	good	yes																																																																										

		13	high	high	low	bad	no	
		14	medium	medium	high	good	yes	
ii.	List down the steps of PCA Using PCA compute the transformed matrix of A Where A is $\begin{bmatrix} 1 & 2 \\ 0.5 & 1.5 \\ 0 & 0.5 \\ -0.5 & 0.25 \end{bmatrix}$							

Q3. (20 Marks Each)	
A	Solve any Two 5 marks each
i.	Define logit function. Explain the importance of logit function in logistic regression with appropriate example
ii.	Given  $X = [3, 5] \quad W = \begin{bmatrix} 1 & 2 \\ 4 & -2 \end{bmatrix} \quad Y = [1, -5] \quad c = 1$ Compute output Z using binary bipolar activation function. Also compute the new weights $y_1, y_2, w_{11}, w_{12}, w_{21}, w_{22}$
iii.	Define covariance ? For the given dataset, compute the covariance matrix

	X_1 X_2 2.5 2.4 0.5 0.7 2.2 2.9 1.9 2.2 3.1 3.0 2.3 2.7 2.0 1.6 1.0 1.1 1.5 1.6 1.2 0.9
B	Solve any One 10 marks each
ii.	Explain Linear Separability problem? (2) Solve a linearly separable problem (AND Gate) Solve a linearly non separable problem (XOR gate) both using McCulloch Pitt Model ?
ii.	What is the role of radial basis function in separating nonlinear patterns? Explain with XOR Example.