

**MACHINE LEARNING  
(INFO 4122)**

**Time Allotted : 3 hrs**

**Full Marks : 70**

*Figures out of the right margin indicate full marks.*

*Candidates are required to answer Group A and any 5 (five) from Group B to E, taking at least one from each group.*

*Candidates are required to give answer in their own words as far as practicable.*

**Group – A  
(Multiple Choice Type Questions)**

1. Choose the correct alternative for the following: **10 × 1 = 10**
- (i) Logistic regression is used when you want to \_\_\_\_\_  
(a) predict a dichotomous variable from continuous or dichotomous variables  
(b) predict a continuous variable from dichotomous variables  
(c) predict any categorical variable from several other categorical variables  
(d) predict a continuous variable from dichotomous or continuous variables.
- (ii) Back propagation is a learning technique that adjusts weights in the neural network by propagating weight changes  
(a) forward from source to sink                      (b) backward from sink to source  
(c) forward from source to hidden nodes      (d) backward from sink to hidden nodes.
- (iii) For a neural network, which one of these structural assumptions is the one that most affects the trade-off between under-fitting and over fitting?  
(a) Number of hidden nodes                      (b) Learning rate  
(c) Initial choice of weights                      (d) Use of constant term unit input.
- (iv) Multi Layer Perceptron can be used as \_\_\_\_\_  
(a) a classifier    (b) an estimator  
(c) none of (a) and (b)                                      (d) both (a) and (b).
- (v) Statement 1: The error surface followed by the gradient descent back propagation algorithm changes if we change the training data.  
Statement 2: Stochastic gradient descent is always a better idea than batch gradient descent.  
(a) Only statement 1 is true                      (b) Only statement 2 is true  
(c) Both are true                                      (d) Both are false.
- (vi) Let's say, a "linear regression" model perfectly fits the training data (training error is zero). Now, which of the following statement is true?  
(a) You will always have test error zero                      (b) You cannot have test error zero  
(c) None of the above                                      (d) Can't say.
- (vii) This technique uses mean and standard deviation scores to transform realvalued attributes.  
(a) Decimal scaling                                      (b) Min-max normalization  
(c) z-score normalization                                      (d) Logarithmic normalization.

- (viii) Suppose your model is overfitting.  
Which of the following is NOT a valid way to try and reduce the overfitting?
- (a) Increase the amount of training data
  - (b) Improve the optimization algorithm being used for error minimization
  - (c) Decrease the model complexity
  - (d) Reduce the noise in the training data.
- (ix) Which of the following layer of CNN controls the overfitting problem?
- (a) Convolutional Layer
  - (b) Hidden Layer
  - (c) Down Sampling Layer
  - (d) Fully Connected Layer.
- (x) Regarding bias and variance, which of the following statements are true? Here 'high' and 'low' are relative to the ideal model.
- (a) Models which overfit have a high bias
  - (b) Models which overfit have a low bias
  - (c) Models which underfit have a high variance
  - (d) Models which underfit have a low variance.

**Group- B**

2. (a) Define Precision and Recall in the context of classification. [(CO1)(Understand/LOCQ)]  
 (b) Consider the following confusion matrix for a two-class data set on which classification has been done:

	Actual class 1	Actual class 2
Placed in class 1	34	26
Placed in class 2	36	44

Evaluate the overall accuracy of this classification results. Also calculate the Precision and Recall values. [(CO1)(Analyze/IOCQ)]

- (c) Distinguish between ID3 and C4.5 decision tree algorithm. [(CO2)(Remember/LOCQ)]  
**4 + 4 + 4 = 12**

3. (a) Explain briefly the difference between input space and feature space. [(CO1)(Remember/LOCQ)]  
 (b) Derive the linear regression formula for multiple dependent variables. [(CO2)(Analyze/LOCQ)]  
 (c) Marks obtained by 10 students in the class test and semester examination in machine learning are provided in the following table. Estimate the marks a student may obtain in the semester examination when she got 20 in class test, using linear regression.

Sl No	Class Test Marks	Semester Marks	Sl No	Class Marks Test	Semester Marks
1	28	63	6	28	51
2	27	49	7	26	66
3	23	43	8	21	36
4	17	36	9	22	31
5	24	39	10	19	37

[(CO2)(Evaluate/HOCQ)]  
**4 + 4 + 4 = 12**

**Group - C**

4. (a) Describe principal component analysis (PCA) in the context of dimensionality reduction. [(C03)(Remember/LOCQ)]  
 (b) Explain the concept of Hierarchical Clustering. [(C02)(Remember/LOCQ)]  
 (c) Describe K-Means clustering algorithms. [(C02)(Remember/LOCQ)]

**4 + 3 + 5 = 12**

5. (a) Explain how kernel function is used in non-linear support vector machines. [(C03)(Understand/LOCQ)]  
 (b) Justify the statement that “One can use infinite-dimensional spaces with the kernel trick” in the perspective of non-linear SVM classification. [(C03)(Analyze/IOCQ)]  
 (c) A linearly separable dataset is given in the following Table. Predict the class of (0.6, 0.8) using a support vector machine classifier. Show all the relevant computations.

X1	X2	Y	Lagrange Multiplier
0.3858	0.4687	+1	65.5261
0.4871	0.611	-1	65.5261
0.9218	0.4103	-1	0
0.7382	0.8936	-1	0
0.1763	0.0579	+1	0
0.4057	0.3529	+1	0
0.9355	0.8132	-1	0
0.2146	0.0099	+1	0

[(C05)(Evaluate/HOCQ)]

**3 + 3 + 6 = 12**

**Group - D**

6. (a) What do you mean by cost function and loss function? [(C01)(Remember/LOCQ)]  
 (b) Why ReLU (Rectified Linear Unit) is most popular activation function? [(C02)(Understand/LOCQ)]  
 (c) Briefly explain the momentum and how is it being incorporated in the back propagation learning technique. [(C02)(Analyze/IOCQ)]

**(2 + 2) + 4 + 4 = 12**

7. (a) Suppose a one-layered neural network with a single weight  $w$  is used to implement a function  $y = 2x + 3 + c$ , where,  $x$  and  $y$  are the input and output parameters respectively, whereas  $c$  is Gaussian noise (random number). Derive the update equation using gradient descent approach to minimize the mean squared error.

[(C05)(Evaluate/HOCQ)]

- (b) Construct the network for the the Boolean function  $x_1 \wedge x_2 \vee (\neg X_3)$  using a multi-layer perceptron. Explain how your network is able to model the said function.

[(C06) (Evaluate/HOCQ)]

**6 + (3 + 3) = 12**

**Group – E**

8. (a) Describe the concept of Radial Basis Function and SoftMax Function. [(CO4)(Remember/LOCQ)]  
 (b) State the differences between Convolution Neural Network and Recurrent Neural Network. [(CO4)(Understand/LOCQ)]  
 (c) Describe the concept of One-Hot encoding and Cross Entropy. [(CO4)(Understand/LOCQ)]  
**4 + 4 + 4 = 12**

9. (a) Given the input matrix and the kernel, perform convolution with stride being 1.

1	0	0
0	0	1
1	1	0
Kernel Matrix		

1	0	1	1	0
0	0	0	1	1
1	0	0	0	1
0	1	1	1	0
1	1	0	1	0
Input Matrix				

[(CO5)(Apply/IOCQ)]

- (b) Apply max-pooling and sum-pooling to the results from the above convolutions. [(CO4)(Apply/IOCQ)]  
**4 + (4 + 4) = 12**

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	53.12	23.96	22.92

**Course Outcome (CO):**

After the completion of the course students will be able to

1. Learn the basics of machine learning paradigm.
2. Understand various machine learning algorithms.
3. Mathematically analyze various machine learning approaches and paradigms
4. Understand the concept of deep learning.
5. Analyze various machine learning techniques to get an insight of when to apply a particular machine learning approach.
6. Apply Machine Learning algorithms in practice and implementing their own using real-world data.

\*LOCQ: Lower Order Cognitive Question; IOCQ: Intermediate Order Cognitive Question; HOCQ: Higher Order Cognitive Question.