

**MACHINE LEARNING
(MCAP 2150)**

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) Which of the following ML algorithm has low bias?
(a) Linear Regression (b) Linear Discriminant Analysis
(c) Support Vector Machine (d) Logistic Regression
- (ii) Application of Machine Learning is
(a) email filtering (b) sentimental analysis
(c) face recognition (d) all of (a), (b) & (c).
- (iii) _____ is the machine learning algorithms that can be used with labeled data.
(a) Regression algorithm (b) Clustering algorithm
(c) Association algorithm (d) All the above.
- (iv) The output of training process in machine learning is_____.
(a) machine learning algorithm (b) machine learning model
(c) null (d) accuracy
- (v) Which of the following is a measure to reduce high variance?
(a) Increase the input features as the model is underfitted
(b) Reduce the input features or number of parameters as a model is overfitted
(c) Decrease the regularization term
(d) Use more complex models, such as including some polynomial features.
- (vi) Fraud Detection, Image Classification, Diagnostic, and Customer Retention are applications in which of the following?
(a) Supervised Learning: Regression (b) Supervised Learning: Classification
(c) Unsupervised Learning: Clustering (d) Reinforcement Learning.
- (vii) The VC dimension of a Perceptron is _____ the VC dimension of a simple linear SVM.
(a) larger than (b) smaller than (c) same as (d) not at all related
- (viii) You are given reviews of few Netflix series marked as positive, negative and neutral. Classifying reviews of a new netflix series is an example of_____
(a) unsupervised learning (b) semi supervised learning
(c) supervised learning (d) reinforcement learning.

- (ix) In multiclass classification number of classes must be_____
- (a) equals to two (b) less than two (c) greater than two (d) none.
- (x) Which of the following can only be used when training data are linearly separable?
- (a) Linear logistic regression (b) Linear soft margin svm
(c) Linear hard-margin svm (d) The centroid method.

Group – B

2. (a) Compare between supervised and unsupervised machine learning with suitable real world examples. [[CO1](Understand/LOCQ)]
(b) Using python libraries develop a code to implement logistic regression over a sample dataset. [[CO2](Create/HOCQ)]
6 + 6 = 12
3. (a) Define Machine Learning. [[CO1](Remember/LOCQ)]
(b) Briefly explain the categories of Machine Learning with suitable examples to each. [[CO1](Understand/LOCQ)]
(c) Explain overfitting and underfitting in regression problem with suitable diagram. [[CO2](Analyze/IOCQ)]
2 + 4 + 6 = 12

Group - C

4. (a) A database contains 80 records on a particular topic. A search was conducted on that topic and 60 records were retrieved. Of the 60 records retrieved, 45 were relevant. Calculate the precision, recall and F1 scores for the search procedure. [[CO3](Understand/LOCQ)]
(b) How cross-validation helps to avoid overfitting? Explain with suitable example. [[CO3](Apply/IOCQ)]
6 + 6 = 12
5. (a) Outline the bias-variance tradeoff for error estimation of a model with suitable example. Use suitable diagram to explain the same. [[CO3](Analyze/IOCQ)]
(b) $h(x) = \begin{cases} 1 & \text{if } a \leq x \leq b \\ 0 & \text{else} \end{cases}$, with parameters $(a, b) \in R^2$
Find the corresponding VC-dimension. [[CO3](Apply/IOCQ)]
6 + 6 = 12

Group – D

6. (a) Explain the main objective of using SVM kernel with an example. Briefly express the difference between logistic regression and SVM. [[CO5] (Remember/LOCQ)]

(b) Consider the following data set :

Outlook			Temperature			Play Tennis	
	Y	N		Y	N		
Sunny	2	3	Hot	2	2	Yes	9
Overcast	4	0	Mild	4	2	No	5
Rainy	3	2	Cool	3	1		

If the new data set Today is {Sunny, Hot}, using Naive Bays classifier determine whether playing Tennis will be held?
[[CO4](Apply/IOCQ)]
6 + 6 = 12

7. (a) “A Bayesian Network is a compact representation of the joint probability distribution”.— Justify with a suitable example. [[CO4](Evaluate/HOCQ)]
(b) Explain how kernel function is used in non-linear support vector machines. Also justify the statement that “One can use infinite-dimensional spaces with the kernel trick” in the perspective of non-linear SVM classification. [[CO5](Explain/LOCQ)]
6 + 6 = 12

Group - E

8. (a) Explain the term perceptron? Briefly describe the reasons for using activation function in Neural network. [[CO6](Remember/LOCQ)]
(b) “Overfitting is one of the most common problems every Machine Learning practitioner faces”. Explain some methods to avoid overfitting in Neural Networks. [[CO6](Explain/LOCQ)]
6 + 6 =12
9. (a) Describe the reasons for preferring CNN than ANN for image dataset as input. [[CO6](Remember/LOCQ)]
(b) Explain different layers of CNN. [[CO6](Explain/LOCQ)]
(c) Explain the significance of the RELU Activation function in Convolution Neural Network. [[CO6](Explain/LOCQ)]
4 + 4 + 4 = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	56.25	31.25	12.5

Course Outcome (CO):

After the completion of the course students will be able to
MCAP2150.1: Review the basics of machine learning paradigm and categorize between different types of learning.
MCAP2150.2: Apply linear or logistic regression for suitable dataset.

MCAP2150.3: Relate the generic Computational Learning Theory with real-time problems and measure the relative strength of different learning models.

MCAP2150.4: Predict data using the Naive Bayes algorithm.

MCAP2150.5: Classify data points using SVM.

MCAP2150.6: Construct neural networks and Deep learning models for classification problems.

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