MACHINE LEARNING TECHNIQUES
(AEIE 4233)

Time Allotted :2½ hrs **Full Marks : 60**

Figures out of the right margin indicate full marks.

Candidates are required to answer Group A and any 4 (four) from Group B to E, taking one from each group.

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

Group – A

1. Answer any twelve: **12 × 1 = 12**

Choose the correct alternative for the following

(i) Choosea disadvantage of decision trees among thefollowings?
(a) Decision trees are robust to outliers (b) Factor analysis
(c) Decision trees are prone to overfit (d) All of the above

(ii) Analysis of ML algorithm needs
(a) Statistical learning theory (b) Computational learning theory
(c) Both (a) and (b) (d) None of the above

(iii) In which of the following learning the teacher returns reward and punishment to learner?
(a) Active learning (b) Reinforce learning
(c) Unsupervised learning (d) Supervised learning

(iv) The correlation for the values of two variables moving in the same direction is?
(a) Positive (b) Negative
(c) Neither positive nor negative (d) No correlation

(v) Which machine learning algorithm can be used with unlabelled data?
(a) Regression algorithm (b) Clustering algorithm
(c) Instance based learning algorithm (d) All of the above

(vi) Which of the followingis used as an input to the machine learning model for training and prediction purpose?
(a) Feature (b) Feature vector
(c) Both (a) and (b) (d) None of the above

(vii) PCA is a
(a) Forward feature selection (b) Backward feature selection
(c) Feature extraction (d) All of these

(viii) In CNN, what is the role of the “padding” in the input image?
(a) Enhance image contrast (b) Reduce the image size
(c) Prevent spatial down sampling (d) All of these

(ix) What does dimensionality reduction reduce?
(a) Stochastics (b) Collinearity (c) Performance (d) Entropy

(x) What is back propagation?
(a) It is another name given to the curvy function in the perceptron
(b) It is the transmission of error back through the network to adjust the inputs
(c) It is the transmission of error back through the network to allow weights to be adjusted so that the network can learn
(d) None of the mentioned

Fill in the blanks with the correct word

(xi) The activation function commonly used in CNN is _____.

(xii) The primary purpose of pooling layers in CNN is to reduce _____.

(xiii) If machine learning model output involves target variable, then it is called_____ model.

(xiv) A measurable property of parameter or the data set is _____.

(xv) In _____ feature selection method, we start with an empty feature set.

Group - B

2. (a) State the main differences between supervised learning and reinforcement learning. Give four example of machine learning. *[(CO1)(Remember/LOCQ)]*
(b) State the supervised machine learning problem and give its mathematical representation. *[(CO1) (Understand/LOCQ)]*
(c) What is loss function? Explain L1-loss, L2-loss and binary cross entropy loss functions with their characteristics and figures. *[(CO1)(Understand/LOCQ)]*

3 + 3 + 6 = 12

3. (a) What is regularization? Explain L1 and L2-regularization in the formulation of objective functions in machine learning algorithms. *[(CO1)(Understand/LOCQ)]*
(b) Explain Bias-Variance Trade-off with diagram? *[(CO1) (Understand/LOCQ)]*
(c) Explain feature selection and feature elimination method for dimensionality reduction. *[(CO1)(Analyze/IOCQ)]*

(2 + 4) + 2 + 4 = 12

Group - C

4. (a) List the differences between classification and regression algorithms. *[(CO2)(Understand/LOCQ)]*
(b) State the multivariate linear regression problem. *[(CO2) (Understand/LOCQ)]*
(c) Consider the example below where the mass, y (grams), of a chemical is related to the time, x (seconds), for which the chemical reaction has been taking place according to the table:

Time (seconds)	5	7	12	16	20
Mass (grams)	40	120	180	210	240

Apply least square algorithm to this data set to find the linear regression line. *[(CO2)(Apply/IOCQ)]*

2 + 3 + 7 = 12

5. (a) What are the characteristics of a high variance model? What are the ways to overcome this? *[(CO2)(Remember/LOCQ)]*

(b) Explain the formulation of LASSO regression problem and state its characteristic. How LASSO differs from Ridge regression problem? *[(CO2) (Understand/LOCQ)]*

(c) Explain the steps of principal component regression. How it helps to improve performance of linear regression model? *[(CO2)(Understand/LOCQ)]*

(2 + 1) + (2 + 2 + 1) + 4 = 12

Group - D

6. (a) State the limitations of KNN and Bayes optimal classifiers. *[(CO3)(Remember/LOCQ)]*

(b) Suppose the data mining task is to cluster the following seven points A1(1,1), A2(1.5,2), A3(3,4), A4(5,7), A5(3.5,5), A6(4.5, 5), A7(3.5, 4.5)(with (x,y) representing location) into two clusters. The distance function is the city block distance. Suppose initially we assign A1 and A5 as the centre for the two clusters, respectively. Using K-means algorithm to find the two clusters and their centres after two rounds of executions? *[(CO4) (Apply/IOCQ)]*

4 + 8 = 12

7. (a) Differentiate between hard margin SVM and soft margin SVM. *[(CO3)(Understand/LOCQ)]*

(b) Let us have a logistic regression model to predict whether a person is male or female based on their height. The learned model coefficient values are b0=-100 and b1=0.6. The model gives an output 0if p(male) < 0.5 and 1 if p(male) >= 0.5. Using the logistic regression model for binary classification calculate whether a person of height of 150 cm or more formally is male or female. Test the same model with a height of 120 cm. *[(CO3) (Apply/IOCQ)]*

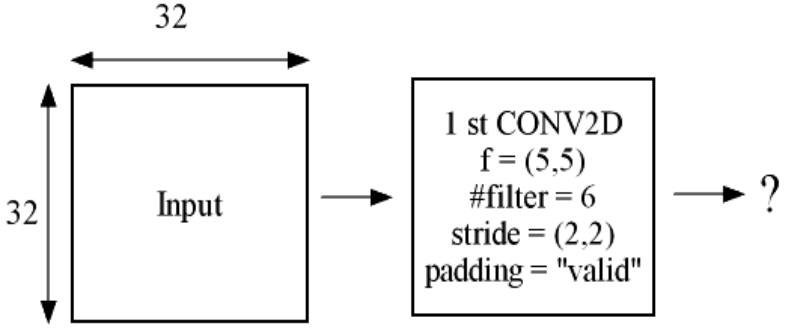
(c) List the importance and limitations of PCA. *[(CO4)(Remember/LOCQ)]*

4 + 6 + 2 = 12

Group - E

8. (a) Suppose you are doing full batch gradient decent using an entire training set (not stochastic gradient decent). Is it necessary to shuffle the training data? Explain your answer. *[(CO5)(Understand/LOCQ)]*

(b) Consider the 2D-convolution operation of a CNN model as shown in figure below:

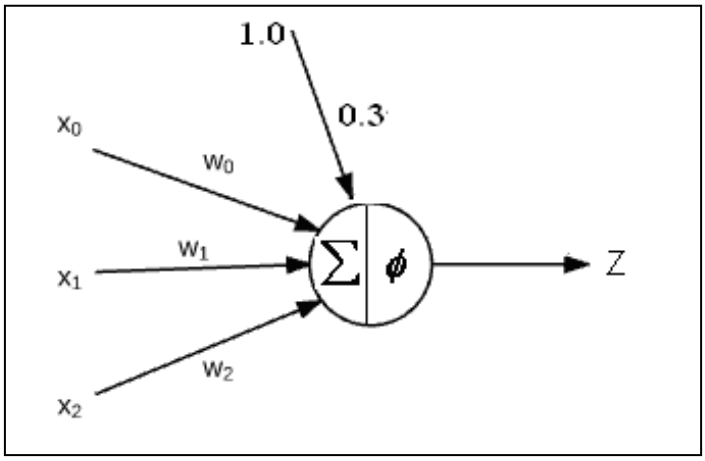


The input shape of the images is$(N_H, N_W, N_C) = (32,32,1)$. Work out the output shape after performing the 1st 2D-convolution operation as shown in the model. *[(CO6) (Apply/IOCQ)]*

(c) Suppose you are given a set of historical images to classify objects into 3 classes: Antiquity, Middle Ages and Modern Era. There are total 5000RGB images. The images are divided after visually inspecting the data set according the image resolution into 64% training, 16% validation and 20% testing set. After this you realize that the training set only contains pictures taken during the day, whereas the validation set only has pictures taken at night. (i) Explain what the issue is and how you would correct it? (ii) As you train your model, you realize that you do not have enough data. Cite 3 data augmentation techniques that can be used to overcome the shortage of data. *[(CO6)(Evaluate/HOCQ)]*

2 + 4 + 6 = 12

9. (a) A single layer neural network with 3 inputs and 1 output with a bias is shown in fig. below. Determine the output (Z) from the network if the activation functions are:
(i) binary sigmoidal and
(ii) bipolar sigmoidal.
The inputs values (x_0, x_1, x_2) are 0.8, 0.6 and 0.4 respectively and their corresponding synaptic weights (w_0, w_1, w_2) are 0.1, 0.3 and -0.2.



[(CO6)(Apply/IOCQ)]

(b) Why does the performance of Deep Learning improve as more data is fed to it? *[(CO5) (Understand/LOCQ)]*

(c) Can CNN be used to perform Dimensionality Reduction? If yes, how? Explain the process of flattering the dimensionality reduction. *[(CO5)(Analyze/IOCQ)]*

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

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	53.12	40.63	6.25

Course Outcome (CO):

After the completion of the course students will be able to

1. Familiarize with the basic concepts and techniques of machine learning and integrate multiple facets of practical machine learning in a single system: data pre-processing, learning, regularization, model selection, data analysis, and hypothesis testing.
2. Develop linear and multivariate regression models on given data and analyze their performance by calculation of R-square and goodness of fit.
3. Implement and analyze existing learning algorithms, including well-studied methods for development of classifier models such as LDA, Bayes, KNN, SVM and logistic regression.
4. Learn data clustering techniques and dimensionality reduction of data by principal component analysis method and apply them on practical problems.
5. Gain knowledge on artificial neural network, convolution neural network and deep learning and implement them with python.
6. Apply basic principles of AI in real world problems that require solving, inference, perception, knowledge representation, and learning

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