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)   Which of the following is/are true for RBF network?
    I.   The network training is divided into two stages: first the input vectors are non-linearly mapped and then linearly classifier is trained.
    II. The training/learning is slow as compared to MLP.
    (a) Only I   (b) Only II   (c) Both I and II   (d) None of these

    (ii)   The minimum time complexity for training an SVM is O(n2). According to this fact, sizes of datasets are not best suited for SVM's are
    (a) Large datasets   (b) Small datasets
    (c) Medium sized datasets   (d) Small and medium sized datasets

    (iii)   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

    (iv)   What happens when you get features in lower dimensions using PCA?
    I.   The features must carry all information present in data.
    II. The features will lose interpretability
    III. The features will still have interpretability
    IV. The features may not carry all information present in data.
    (a) II and IV   (b) I and IV
    (c) II and III   (d) All of these

    (v)   The input image has been converted into a matrix of size 28×28 and a kernel/filter of size 7×7 with a stride of 1. What will be the size of convoluted matrix?
    (a) 22×22   (b) 21×21
    (c) 28×28   (d) 7×7

(vi)   The generalization error in terms of support vector machine is
        (a) How far the hyperplane is from the support vectors
        (b) How accurately the SVM can predict outcomes for unseen data
        (c) The threshold amount of error in an SVM
        (d) Both (b) and (a)

(vii)   When zero-padding is not added, it is called
        (a) Valid Padding   (b) No-Padding
        (c) Wide Padding   (d) Narrow Convolution

(viii) Signal transmission at synapse is a __
        (a) physical process   (b) chemical process
        (c) physical & chemical both   (d) electrical

(ix)   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.

(x)   What will be the output of the convolution of the input (I) and kernel (g)? Consider I and g as given bellow.

$$I = \begin{matrix} 0.0 & 1.0 & 2.0 \\ 0.1 & 1.1 & 2.1 \\ 0.2 & 1.2 & 2.2 \end{matrix} \quad g = \begin{matrix} 0 & 0.5 & 0 \\ 0 & 1.0 & 0 \\ 0 & 0.5 & 0 \end{matrix}$$

(a) 3.3   (b) 2.2   (c) 2.0   (d) 2.5

Group- B

2. (a) Explain the purpose of linear regression to develop a data driven model. Suppose your model is $Y = \beta_0 + \beta_1 X$, where X is input (regress) variable and β's are regression coefficients. Y is noted as predictor variable. Show that the regression coefficients are computed using the expression, $\beta = (X^T X)^{-1} X^T Y$.
        [(CO5) (Apply/IOCQ)]

    (b) Distinguish between linear regression and logistic regression.
        [(CO2) (Understand/LOCQ)]

    (c) A coin is tossed 100 times and lands heads 62 times. What is the maximum likelihood estimate for θ the probability of heads.   [(CO2) (Apply/IOCQ)]

    (d) Apply least square method to fit a model $Y = \beta_0 + \beta_1 X$,through the following data points

X	1	2	3	4	5	6	7
Y	9	8	10	12	11	13	15

        [(CO6)(Apply/IOCQ)]
        (2 + 3) + 2 + 2 + 3 = 12

3. (a)   Consider the following set of training examples:

A1	A2	Classification
T	T	+
T	T	+
T	F	-
F	F	+
F	T	-
F	T	-

(i) What is the entropy of this collection of training example with respect to the target function classification?

(ii) What is the information gain of A2 relative to these training examples?

[(CO3) (Apply/IOCQ)]

(b)   "ID3 may suffer from overfitting problem" – justify.   [(CO5) (Evaluate/HOCQ)]

(c)   Distinguish between ID3 and C4.5 decision tree algorithm.

[(CO5)(Analyse/IOCQ)]

(3 + 3) + 3 + 3 = 12

Group - C

4. (a)   What do you mean by dimension reduction?   [(CO1) (Remember/LOCQ)]

(b)   Let the patterns (1, 2), (2, 3), (3, 3), (4, 5), (5, 5) belong to C1 and (1, 0), (2, 1), (3, 1), (3, 2), (5, 3), (6, 5) belong to C2. Now reduce the dimension using Fisher's Linear Discriminant Analysis (FDA).   [(CO6) (Apply/IOCQ)]

1 + 11 = 12

5. (a)   What is the intuition of a large margin classifier ?[(CO1) (Remember/LOCQ)]

(b)   What is a kernel in SVM? Why do we use kernels in SVM? Can we apply kernel trick in logisticregression?   [(CO1) (Remember/LOCQ)]

(c)   Is SVM sensitive to the feature scaling. Justify your answer with example.

[(CO3) (Evaluate/HOCQ)]

(d)   Explain soft margin and hard margin Support Vector Machine?

[(CO2)(Understand/LOCQ)]

2 + (1 + 2 + 1) + 3 + 3 = 12

Group - D

6. (a)   Explain different types of Gradient Descent in detail?[(CO1) (Remember/LOCQ)]

(b)   What is the difference between forward and backward propagation in neural network model?   [(CO2) (Understand/LOCQ)]

(c)   How does the learning rate affect the training in neural network?

[(CO5)(Analyze/IOCQ)]

(d)　　Are neural networks helpful in medicine? Justify your answer with reason.
[(CO3) (Evaluate/HOCQ)]

(e)　　What is the role of activation function in Artificial Neural Network?
[(CO2)(Understand/LOCQ)]
3 + 2 + 3 + 2 + 2 = 12

7.　(a)　　What do you mean by optimizers?　[(CO1) (Remember/LOCQ)]

(b)　　Why zero initialization is not good initialization technique?
[(CO2) (Understand/LOCQ)

(c)　　What do you mean by Backpropagation?　[(CO1)(Remember/LOCQ)]

(d)　　What do you mean by cost function and loss function?
[(CO1) (Remember/LOCQ)]

(e)　　Why RelU (Rectified Linear Unit) is most popular activation function?
[(CO2)(Understand/LOCQ)]
3 + 2 + 3 + 2 + 2 = 12

Group – E

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

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

1	0	0
0	0	1
1	1	0

Kernel 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

9.　(a)　　Justify how a CNN can identify an object even when it is transformed.
[(CO4) (Evaluate/HOCQ)]

(b)　　Explain the architecture of AlexNet deep convolutional network.
[(CO4) (Understand/LOCQ)]
4 + 8 = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	40.63%	46.88%	11.46%

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

Department & Section	Submission Link
IT	https://classroom.google.com/c/NDAwODkyNDgwNDY1/a/NDYzNjcyOTAzNDIw/details