Master of Computer Applications
MCAC 204: Artificial Intelligence and Machine Learning
Unique Paper Code: 223421208
Semester II
Year of admission: 2024
May-2025

Time: Three Hours **Max. Marks: 70**

Instructions:

1. All questions carry equal marks.
2. Use proper notation and show complete working for full credit.
3. Use of a simple calculator is allowed.

1. a. Consider a user-item rating system represented by the following partially observed matrix: **[6]**

$$Y = \begin{bmatrix} 3 & 7 & ? \\ 4 & ? & 6 \end{bmatrix}$$

We want to approximate this matrix using matrix factorization $X = UV^T$ closely approximates the observed ratings in Y. Assume we start by fixing V to the initial values of $[3 \quad 1 \quad 2]^T$. Find the optimal 2×1 vector U in this case. (Express your answer in terms of λ: regularization parameter)

b. Consider the following Convolutional Neural Network (CNN) applied to a database of images of handwritten digits to classify them into one of the digits 0, 1, 2, ..., 9. Each image is of size $28 \times 28 \times 3$. All convolution layers (CONV) have a stride of one with valid padding, and all pooling functions (MAX-POOL) use a stride of 2. The network uses two fully connected layers (FCs). The output layer uses the Softmax function. Fill in the blanks in the following table: (Show the calculations at each step) **[8]**

	Activation Shape	Activation Size	# of trainable parameters
Input	(28, 28, 3)		
CONV1 (f=3)			
MAX-POOL (2×2)			
CONV2 (f=3)			
MAX-POOL (2×2)			
FC1	(120, 1)		
FC2	(84, 1)		
Output	(10, 1)	10	

2. a. For each of the following applications, determine whether precision or recall is the more critical evaluation metric. Provide a brief justification for each case: [1.5+1.5=3]

   i. Fake news detection on social media platforms.
   ii. Identifying faulty components in a manufacturing pipeline.

   b. You are provided with a binary classification dataset containing 1000 samples, each with 8 categorical features. You train an AdaBoost ensemble using decision stumps as the base (weak) classifiers. At the beginning, all training samples are assigned equal weights. The first decision stump is trained and achieves a weighted classification error of 0.2. After the first iteration, the sample weights are updated based on the performance of this stump. [2+2=4]

   i. Compute the Amount of Say (α) assigned to this weak learner in the ensemble.
   ii. Determine the maximum sample weight assigned to any individual sample after the first weight update.

   Assume the AdaBoost algorithm uses the standard exponential update rule, and the base of the logarithm is natural (i.e., ln).

   c. Given a feature matrix $X \in \mathbb{R}^{(m \times n)}$ and a label vector $y \in \{0,1\}^m$ representing a binary classification problem, answer the following: [2+3+2=7]

   i. Derive the cost function used in logistic regression.
   ii. Derive the gradient of this cost function with respect to the model parameters.
   iii. Write the gradient descent update rule used to minimize the cost function.

   Assume the hypothesis function is defined using sigmoid activation.

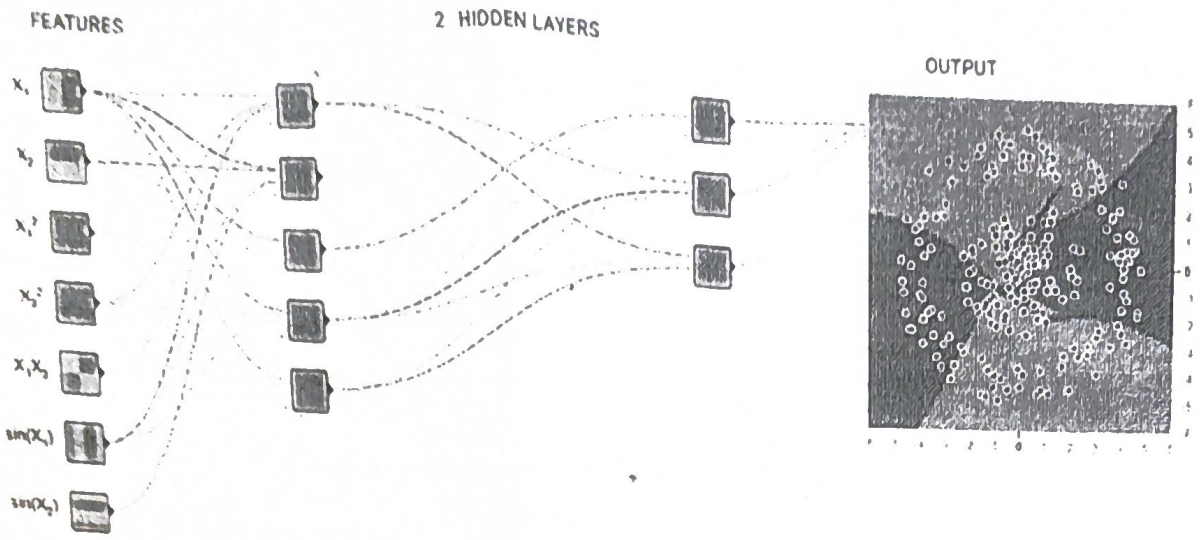
3. a. For a neural network that outputs the likelihood of occurrence of a disease out of six, which activation functions should be employed: Sigmoid, ReLU, SoftMax, and Tanh? Justify your answer. [2]

   b. With the help of a suitable diagram, prove that the width of the margin in the linear support vector classifier (LSVM) is $\frac{2}{\|\vec{w}\|}$, where $\vec{w}$ is a vector perpendicular to the decision boundary. [6]

   c. Find one principal component of the dataset given below (Show all the intermediate steps): [6]

Age	Weight
20	65
25	60
5	15
10	20

4. a. Consider a three-layer neural network with the following configuration: [3+3=6]



i. Write the forward propagation equations for this neural network. Assume that the hidden layers use the ReLU activation function, and the output layer uses the sigmoid activation function.

ii. Calculate the total number of trainable parameters in the network.

b. Consider the following implementation of a neural network: [8]

```
import tensorflow as tf
from tensorflow.keras import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.losses import
BinaryCrossentropy

model = Sequential([
          Dense (units = 10, activation = 'tanh'),
            Dense (units = 5, activation = 'tanh'),
          Dense (units = 1, activation = 'sigmoid')
              )]
model.compile(loss = BinaryCrossentropy())
```

Derive the expressions for the computations of gradients $\frac{\partial L}{\partial w^{[l]}}$ and $\frac{\partial L}{\partial b^{[l]}}$ at each layer l.

5. a. Explain the role of padding in a convolutional layer. What is the difference between "valid" and "same" padding? [3]

b. How do the exploding and vanishing gradient problems affect the training of deep neural networks, and what techniques can be used to mitigate these issues? [3]

c. Consider a training set comprising 5,000,000 records. Write an algorithm with forward and backward propagation steps of the mini-batch gradient descent method with 5000 mini-batches of size 1000 each. [4]

d. Consider the following training set for a two-class classification problem. [4]

Record	A	B	C	Class
1	0	0	0	+
2	0	0	1	−
3	0	1	1	−
4	0	1	1	−
5	0	0	1	+
6	1	0	1	+
7	1	0	1	−
8	1	0	1	+

Use Naïve Bayes classifier to classify a new instance: $<A = 0, B = 1, C = 0>$.

(c) Find two principal components of the dataset given below (Show all the intermediate steps):

3·2 4·2 2·6

X1	X2	X3
4	5	3
2	3	2
6	7	4
1	2	1
3	4	3