Full Marks: 100   **Time: 3 hours**

GROUP A (20 Marks)

Answer the following questions. Each question is of 2 marks.

1.

i) For the following boolean function, the inequalities are given as follows:

X_1	X_2	Output	Inequalities
0	0	0	$w0 + w1 \cdot x1 + w2 \cdot x2 < 0$
1	0	1	$w0 + w1 \cdot x1 + w2 \cdot x2 \geq 0$
0	1	1	$w0 + w1 \cdot x1 + w2 \cdot x2 \geq 0$
1	1	0	$w0 + w1 \cdot x1 + w2 \cdot x2 < 0$

   Calculates the solution values for (w0,w1,w2), so that the above inequalities can be satisfied?

ii) Consider the following table, where x1 and x2 are the inputs, h1,h2,h3 and h4 are the hidden

layer outputs. A perceptron will fire if $\sum_{i=1}^{4} w_i h_i \geq w_0$,w$_i$ represents the hidden layer weights,

   i=1 to 4 and w$_0$ is the bias.

X_1	X_2	$Output$	h_1	h_2	h_3	h_4	$\sum_{i=1}^{4} w_i h_i$
0	0	1	1	0	0	0	w'_1
0	1	0	0	1	0	0	w'_2
1	0	0	0	0	1	0	w'_3
1	1	1	0	0	0	1	w'_4

Derive the inequalities for the above mention Boolean function using perceptron.

iii) For a given data, D={xi,yi}i=1to n , where xi,yi∈R (set of real numbers). Which of the following function(s) can be used to approximate the relation between x and y such that our learning
   algorithm can be trained very well? Specify your answer with proper reason.

a. $\hat{y} = \dfrac{1}{1+e^{-w^T x}}$
b. $\hat{y} = x^T W x$
c. $\hat{y} = \dfrac{1}{1+e^{-x}}$
d.

combination (specify that)

iv) What will be the values of bias and weights when you want to implement 2-input NAND gate using perceptron? (put the values in order of b, w1, w2)

v) Suppose you want to implement 2-input NOR gate. Find the number of error value given w0 = 1, w1=-1, w2=-1? Also specify the inputs related to the error.

vi) Signify the importance of a property named differentiable in deep learning.

vii) Why analytical method of optimization is not applicable in a problem where an explicit expression of objective function is not available?

viii) Why non-convex optimization problems may not be solved by using gradient descent?

ix) Input to a perceptron is X = $[1\ 2\ 1\ 0]^T$ and parameter/weight vector is $W = [0.1\ 0.2\ 0.1\ 0.5]^T$. In case of a binary classification problem, in which class (positive/negative) the above-mentioned input vector will be classified? Consider step activation function

x) If learning rate is set to zero, what will be the impact in gradient descent algorithm?

GROUP B (30 Marks)

Answer the following questions. Each question is of 5 marks.

2. i) Consider the following objective function $J(W) = w_1{}^2 + w_2{}^2 + 3$. Use gradient descent algorithm with learning rate 0.5 for two iterations.

3. i) The Harley–Davidson Iron 883 has the following normalized features:
   (Displacement, Mileage, Kerb Weight, IsRedAvailable) = (8.83,0.20,2.56,1).
   Now consider a person who wants to decide whether to buy a Harley-Davidson Iron 883. He assigns the following weights to each of these inputs: W = [0.9, 0.4, 0.7, 1]. Further, suppose that θ = 8. Based on the above information, do you think he will buy the bike based on a simple McCulloch–Pitts neuron.
   ii) Define linear decision boundary.

4. i) Consider the 4 data points with 3 dimensions that are divided into two classes, Class 1 and Class 2.
   Let the initial weight vector be [–0.5, 1, 0.2]. Apply the simple trial and error-based perceptron learning algorithm to the sample data in the same order. Find the values of the weight
   Vector after the algorithm converges. If required, repeat accessing the samples in the same order till the algorithm converges. (Hint: if xε1, perform w+x, else perform w–x. Continue this
   for every x and redo till the algorithm converge.
   ii) Define probability mass function.

5. A.
  i) Explain the working approach of a feed forward neural network specifying the properties and dimensions of different component.
  ii) Define the relation between scalar, vector, matrix and tensor.
OR
 B.
  i) Input to a perceptron is $X = [2\ 1\ 0\ 1\ 2\ 1\ 0]^T$ and parameter/weight vector is $W = [0.1\ 0.2\ 0.1\ 0.5\ 0.4\ 0.6\ 0.2]^T$. In case of a binary classification problem, find the probability of the membership the above-mentioned input vector in both classes (positive/negative)? Consider sigmoid activation function

6. A.
  i) Discuss about the need and types of activation function in deep learning.
  ii) Define one hot encoding.
OR
 B.
  i) What do you mean by a decision boundary?
  ii) What type of decision boundaries can be learned by Perceptrons?

7.A.
  i) Consider the XNOR operation. Assuming that it is a classification problem with the output being the two classes '0' and '1', check if it is a linearly separable problem.
  ii) Define conditional probability.
 OR
 B.
  i) What is Bias unit of a Perceptron?
  ii) Is it possible to learn all types of linear decision boundary without using bias unit? Justify

GROUP C (50 Marks)

Answer the following questions. Each question is of 10 marks.
8. i) a. Specify the reasons behind the rise of deep learning over machine learning.
    b. Discuss about different types of machine learning algorithm.
  ii) a. How can we use bayes rule in detection of spam mail?
    b. Specify the properties of a biological neuron with a neat diagram.

9. i) Write down the gradient descent algorithm to minimize a function?
  ii) What are the drawbacks of gradient descent algorithm?

10. A.
  i) a. What is linear decision boundary? Specify it's role in learning.
    b. Implement a three input OR using perceptron.
  ii) a. Specify the rule of convergence. Also mention how does it ensure the minimization of error.
    b. What is the purpose of using Bayesian network?
 OR
 B.

i) a. Compare and contrast perceptron over sigmoid neuron.
   b. Point out the reasons behind the rise of perceptron over mc-culloch pit neuron.
ii) a. Write the perceptron learning algorithm and explain briefly how it works.
   b. Discuss about the role of bias in deep learning.

11. A.
   i) a. Define the XOR problem and discuss about the solution.
      b. Discuss about the different activation function of output layer in feed forward neural network.
   ii) a. How do we calculate marginal probability from joint probability distribution? (Use example)
      b. Discuss about universal approximation theorem.

   OR

   B.
   i) What is activation function? Why is it required in ANNs?
   ii) Consider the following objective function $J(W) = w_1{}^3 + w_2{}^2 - 6$. Use gradient descent algorithm with learning rate 0.1 for 4 iterations.

12. A.
   i) a. Establish the relation between loss and learning algorithm using proper example.
      b. Specify the need of matrix factorization and specify a method to do that.
   ii) a. What is dimensionality reduction? Discuss a technique.
      b. What is the alternative efficient approach of replacing loop in implementing complex operation? Explain using example.

   OR

   B.
   i) Why loss functions are used in perceptron training? Write down the sum squared error function?
   ii) Derive the weight update rule of perceptron learning with sum squared error function?

****