

- NB:
- (1) Question No. 1 is compulsory.
 - (2) Attempt any three questions out of the remaining five.
 - (3) All questions carry equal marks.
 - (4) Assume suitable data, if required and state it clearly.

ATTEMPT ANY FOUR.

- Q1. a. Design AND gate using Perceptron. [20]
 b. Suppose we have N input-output pairs. Our goal is to find the parameters that predict the output y from the input x according to some function $y = x^w$. Calculate the sum-of squared error function E between predictions y and inputs x . The parameter w can be determined iteratively using gradient descent. For the calculated error function E , derive the gradient descent update rule $w \leftarrow w - \alpha \frac{dE}{dw}$.
 c. Explain dropout. How does it solve the problem of overfitting?
 d. Explain denoising auto encoder model.
 e. Describe sequence learning problem.
- Q2. a. Explain Gated Recurrent Unit in detail. [10]
 b. What is an activation function? Describe any four activation functions. [10]
- Q3. a. Explain CNN architecture in detail. Suppose, we have input volume of $32 \times 32 \times 3$ for a layer in CNN and there are ten 5×5 filters with stride 1 and pad 2; calculate the number of parameters in this layer of CNN. [10]
 b. Explain early stopping, batch normalization, and data augmentation. [10]
- Q4. a. Explain RNN architecture in detail. [10]
 b. Explain the working of Generative Adversarial Network. [10]
- Q5. a. Explain Stochastic Gradient Descent and momentum based gradient descent optimization techniques. [10]
 b. Explain LSTM architecture. [10]
- Q6. a. Describe LeNET architecture. [10]
 b. Explain vanishing and exploding gradient in RNNs. [10]
