2211N815

Candidate's Seat No :_____

IM.Sc AIML Sem.-7 Examination

CC 403

Deep Learning

Time : 2-30 Hours]

November-2024

[Max. Marks : 70

Instructions: All questions are compulsory. Use of non-programmable scientific calculator is allowed.

Q.1 **(a)** Describe the characteristics of Feedforward Neural Networks (FNN) along with its applications. **(07)**

(b) Why is linear activation not suitable for classification problems? How do activation functions differ in classification vs. regression models? **(07)**

OR

(a) Briefly describe advantages and disadvantages of artificial neural network (ANN). **(07)**

(b) Why is it a bad idea to insert a hidden layer with a single neuron? What could go wrong? **(07)**

Q.2 **(a)** What would happen if no activation functions were used in a neural network? Explain with example. **(07)**

(b) Describe the backpropagation algorithm and its significance in training ANNs. **(07)**

OR

(a) What are the key differences between RMSProp and Adam in terms of optimization? **(07)**

(b) Provide an example where the gradients vanish for the sigmoid activation function. **(07)**

Q.3 **(a)** Suppose you have a single parameter θ and you are using the AdaGrad optimization algorithm to minimize a loss function $L(\theta)$. The gradient of the loss function with respect to θ at the current time step t is $g_t = 0.03$. The initial learning rate $\eta = 0.01$ and $\varepsilon = 10^{-8}$. Assume that the sum of squared gradients up to time step $t-1$ is $G_{t-1} = 0.02$. **(07)**

(a) Calculate the updated sum of squared gradients G_t at time t.

(b) Determine the updated parameter θ_{t+1} using AdaGrad.

(b) State clearly key differences between L1 and L2 Regularization. **(07)**

OR

(a) What distinguishes VGGNet from other architectures? **(07)**

(b) Compare the performance of Adagrad and Adam in terms of convergence speed and stability. **(07)**

Q.4 **(a)** Discuss the working principle of recurrent neural networks, and explain how they can be used to process sequential data. **(07)**

(b) What is the vanishing gradient problem in RNNs? **(07)**

OR

1

(P.T.O)

(a) Discus the working principle of LSTM network. (07)

(b) Illustrate the working of Dropout in neural network. (07)

Q.5 Attempt any **SEVEN** out of **TWELVE:** (14)

(1) What is the key difference in the activation functions used in the perceptron and logistic regression models?

(2) What is the role of bias in an artificial neuron, and how would removing the bias term affect the learning capacity of a neural network?

(3) Compute the total number parameters (Weights and bias) when Layer sizes: 10, 16, 16, 16, 2.

(4) Given an input image of size 32x32x3 (height x width x channels) and a filter of size 5x5x3 with a stride of 1 and no padding, what will be the dimensions of the output feature map?

(5) Calculate the number of parameters in a convolutional layer with 10 filters, each of size 3x3x3, and an input image of size 64x64x3.

(6) For an input image of size 64x64x3, if you apply a max-pooling layer with a filter size of 2x2 and a stride of 2, what will be the dimensions of the output feature map?

(7) How does the learning rate in Adam differ from that in traditional stochastic gradient descent (SGD)?

(8) In what scenario would you choose RMSProp instead of Adam for training a deep learning model?

(9) What is the purpose of using kernels or filters in convolution?

(10) What are the key components of an LSTM cell that help it retain information over long sequences?

(11) Why does logistic regression use a log-likelihood loss?

(12) What distinguishes a Gated Recurrent Unit (GRU) from a Long Short-Term Memory (LSTM) network?

**** ****