Ref. No. : Ex/CSE/PE/B/T/413C/2024

B.E. COMPUTER SCIENCE AND ENGINEERING FOURTH YEAR FIRST SEMESTER - 2024

DEEP LEARNING

Time: 3 hours | Full Marks: 100

Answer Question No. **1 (Compulsory)** and answer **FOUR** questions by choosing ONE from each group

1.

a. What are the basic differences between traditional programming and machine learning? -Explain.
b. Explain the advantages of the Logistic Regression Model over the threshold-based classification model.
c. Derive the weight update rule used for updating weights of a regression model.
d. What is momentum? How can it affect the training performance of an ANN model. Can you relate it to physics?
e. Is CNN rotation invariant by nature? Explain how this issue can be resolved to some extent using multiple filters.

4 x 5 = **20 marks**

GROUP A

(Answer Question No. 2 **or** Question No. 3)

2. a) Differentiate (with examples) among supervised, unsupervised and semi-supervised learning .
b) In deep learning, what does "depth" refer to? Explain with an example.
c) Consider a dataset in which two data points {(-1, -1), (1, 1)} belongs to one class, and the other two points {(1,-1), (-1, 1)} belong to the other class . Start with weight parameter values of a linear neuron model at (0.5,0.5), and use *perceptron rule* for weight updates (incremental mode). Set *learning rate*: $\alpha = 1$ and the bias weight = 0.5. Show calculations up to 2 epochs

(5+5+10) = **20 marks**

3. a) For a multivariate linear regression model, if X is an N x p data matrix with an p-dimensional input vector per row and y is an N-dimensional vector of outputs for the training examples, then show that

$$w^* = (X^T X)^{-1} X^T y$$

where w^* is the best vector of weights.

b) Considering L2 loss function derive the weight update rule for the logistic regression model.

(10 + 10) = **20 marks**

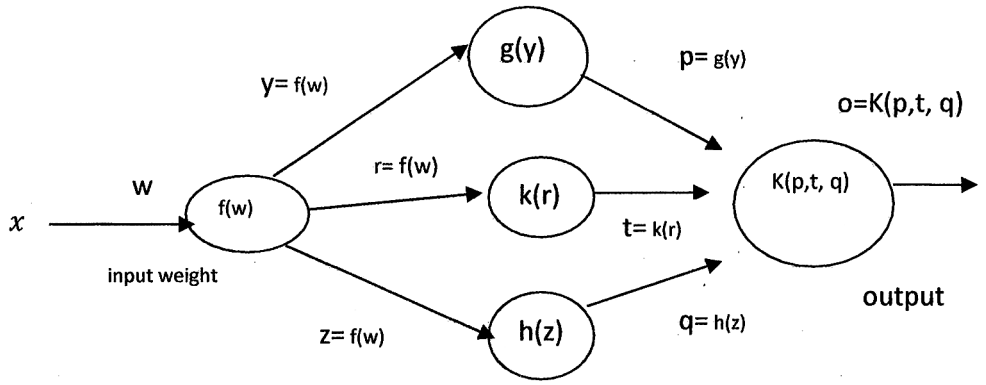
[Turn over

GROUP B

(Answer Question no. 4 **or** Question no. 5)

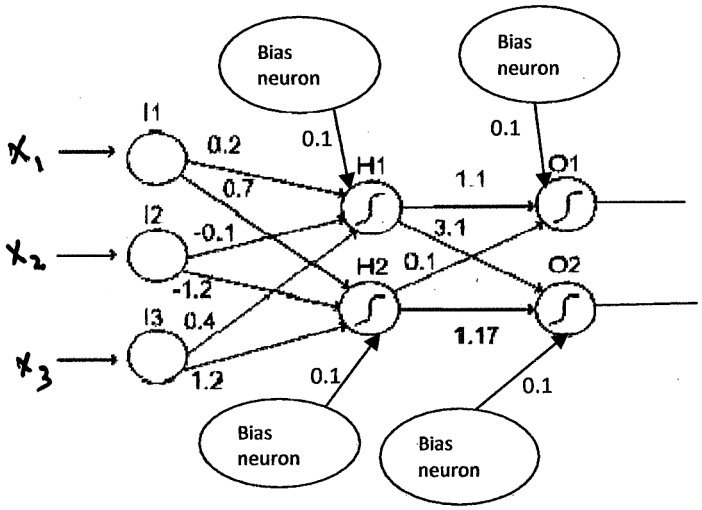
4. a) What is called regularization? Explain a regularization method used in Artificial Neural Networks .

b) Consider the following computational graph and find $\frac{\delta o}{\delta w}$ using chain rule of calculus.



c) Derive the input to hidden layer connection weight update rule for ANN when the hidden nodes use *Relu* activation functions and the output nodes use sigmoid activation functions. Consider $\boldsymbol{Relu'(x) = 0\ if\ x < 0, else\ Relu'(x) = 1}$. (5+ 5+10) =**20 marks**

5. The following is an Artificial Neural Networks, with sigmoid units in the hidden layer and the output layers. The weights have been set arbitrarily between all the units.



Consider that the following training example is submitted to the net (shown in the above figure).

X1	X2	X3	O1	O2
0.2	0.5	0.3	**0.9**	**0.1**

Now show the forward pass to compute responses at H1, H2 and O1, O2. Then use backpropagation algorithm to find updates for weights associated with the connections between hidden and output layers. Assume the learning rate η = 0.3 and momentum = 0.5. You may assume the values of the other hyperparameters if necessary. 8 + 12 = **20 marks**

GROUP C

(Answer Question no. 6 or Question no. 7)

6. a) Explain the Adam optimization algorithm.
   b) What is skip connection? Why is it used?
   c) In CNN, what is the difference between max pooling and global ~~max~~ mean pooling. Give diagrams in support of your answer

10+ 5+5 = **20 marks**

7) a) What do you by exponentially weighted average? Explain with an example.
   Compare Gradient descent with momentum and RMSprop optimization algorithm.
   b) Explain with a diagram how weight sharing is done in CNN.
   c) What is batch normalization? Why is it used?

(5+5) +5+5 = **20 marks**

GROUP D

(Answer Question no. 8 **or** Question no. 9)

8. a) What is the drawback of the RNN network ? How is it overcome in the LSTM network?
b) Write the mathematical equation for the cumulative internal state of an LSTM unit. In this equation, if the input gate is close to 0 and the forget gate is close to 1, then what will be transferred to the next state. What is its significance?
c) Draw a detailed architecture of a CNN model used for the handwritten digit recognition task. Show clearly the dimensions at each layer. Consider the following configuration of the network:

(i) Input image size is (32 x 32)
(ii) The first convolution layer uses 6 filters with the filter size (5 x 5) and stride = 1. The size of the max pooling layer added to this Conv. layer is (2 x 2) with stride = 2.
(iv) The first fully connected layer with 100 hidden nodes and the second fully connected layer with 80 hidden nodes.
(v) The last fully connected layer with m hidden nodes where m is to be determined by analyzing the given problem.

d) Consider the following model definition written in Keras code for implementing CNN model for image classification task and answer the following questions.

```
model = Sequential()
model.add(Conv2D(50, kernel_size=(3,3), strides=(1,1), padding='same',
activation='relu', input_shape=(32, 32, 3)))
model.add(Conv2D(75, kernel_size=(3,3), strides=(1,1), padding='same',
activation='relu'))
model.add(MaxPool2D(pool_size=(2,2)))
model.add(Dropout(0.25))
model.add(Conv2D(125, kernel_size=(3,3), strides=(1,1), padding='same',
activation='relu'))
model.add(MaxPool2D(pool_size=(2,2)))
model.add(Dropout(0.25))
model.add(Flatten())
model.add(Dense(500, activation='relu'))
model.add(Dropout(0.4))
model.add(Dense(250, activation='relu'))
model.add(Dropout(0.3))
model.add(Dense(10, activation='softmax'))
```

(i) In the above code, *ReLu activations* have been used in the intermediate layers, but *softmax activation has* been used at the output layer - Why?.
(ii) In the above code, what is the purpose of Flatten()?- Explain. (5 x 4) = **20 marks**

9. a) What is the purpose of Dropout? Explain briefly how Dropout is implemented in a Deep Neural Network.

b) Write the mathematical equations used for implementing the following in an LSTM unit:
(i) Effect of input (ii) Cumulative internal state (iii) Overall output value
From the equation of the cumulative internal state , derive what happens If the input gate is close to 1 and the forget gate is close to 0? What is its significance? - explain.

c) Consider the Keras code given above in question no. 8(d) and answer the following questions

i) In the code "`model.add(Conv2D(50, kernel_size=(3,3), strides=(1,1), padding='same', activation='relu', input_shape=(32, 32, 3)))`", what are meanings of the values of the arguments used in Conv2D( )?

(ii) What will be the dimension of the feature map produced after applying the first convolutional layer? 5+ (6+4) +5=20 marks

---