Name of the Examinations: B.E. COMPUTER SCIENCE AND ENGINEERING FOURTH YEAR FIRST SEMESTER SUPPLEMENTARY EXAM - 2024

Subject : 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 supervised and unsupervised learning? -Explain with examples.

b) What are the disadvantages of the Logistic Regression Model with MSE (Mean Square Error) loss? How can it be overcome?

c) What are drawbacks of a threshold-based classifier? -explain.

d) Write the weight update rules used in the BP algorithm with momentum. What is the use of momentum? -explain.

e) Is CNN translation invariant by nature? Explain your answer with justification.

4 x 5 = **20 marks**

GROUP A

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

2. 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 MSE loss function, derive the weight update rules used for updating weights of a linear regression model. (Use Gradient descent Search)

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

3. a) What is semi-supervised learning ? Give an example with justification.

b) What are differences between traditional machine learning and deep learning?

c) Consider a dataset in which two data points (1, 1) belongs to one class, and the other three data points {(-1, -1), (1,-1), (-1, 1)} belong to the other class . Start with weight parameter values of a linear neuron model at (0.2, 0.2), and use *perceptron rule* for weight updates (incremental mode). Set *learning rate*: $\alpha = 1$ and the bias weight = 0.2. Show calculations up to 2 epochs

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

[Turn over

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

4. a) Describe the structure of an artificial neuron? How it is similar to a biological neuron?

   b) Explain the basic structure of multilayer perceptron model . Explain how it can be used to solve XOR problem.

   c) Derive the input to hidden layer connection weight update rule for ANN when the hidden nodes use sigmoid activation functions and the output nodes also use sigmoid activation functions.

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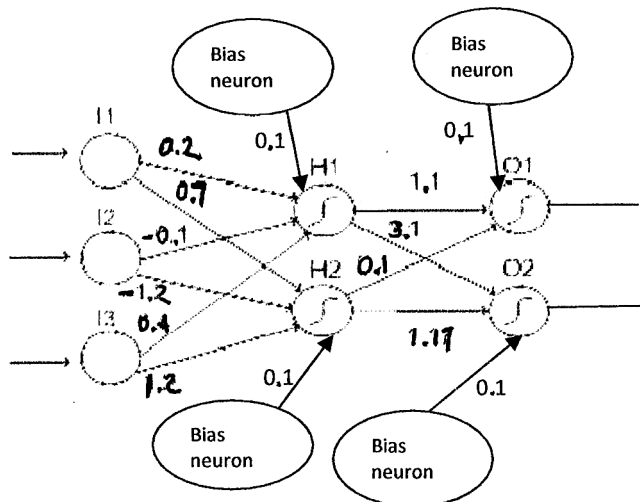
Consider that the following training example is submitted to the net (shown in the above figure).

X1	X2	X3	O1	O2
0.3	0.2	0.7	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 = 0.2. 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 RMSprop optimization algorithm.

   b) Draw a single residual block with skip connection and explain the use of skip connection.

   c) In CNN, what are the functions of convolution layer and max pooling layer - explain.

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

7) a) What do you mean by exponentially weighted average? Explain it with an example. Highlight the differences between Adam optimization algorithm and RMSprop optimization algorithm.

   b) Compare traditional ANN and CNN in the context of image classification task.

   c) What is batch normalization? When is it useful?

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

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

8. a) Compare and contrast RNN network and 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 1 and the forget gate is close to 0, 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 English character (consider 26 alphabets only) recognition task. Show clearly the dimensions at each layer. Consider the following configuration of the network:

   (i) Input image size is (64 x 64)

   (ii) The first convolution layer uses 12 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 200 hidden nodes and the second fully connected layer with 160 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(25, kernel_size=(3,3), strides=(1,1), padding='same',
activation='relu', input_shape=(32, 32, 3)))
model.add(Conv2D(50, 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(100, 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(300, activation='relu'))
model.add(Dropout(0.4))
model.add(Dense(150, activation='relu'))
model.add(Dropout(0.3))
model.add(Dense(10, activation='softmax'))
```

(i) In the above code, `Dropout(0.25) and Dropout(0.4))` have been used. What is the use of droprout in this cases? What do the arguments of Dropout function, 0.25 and 0.4 signify? Why is *softmax activation* used at the last layer and what is its purpose?

(ii) In the above code, what is the function of dense layers?- Explain. If no dense layer is used, will this CNN model work? -Explain.   (5 x 4) = **20 marks**

9. a) What is gradient vanishing and gradient explosion problem? Explain.

   b) Draw a diagram of an LSTM unit depicting the three gates used in it. Explain the functions of these three gates.

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

   i) In the code " `model.add(Dense(10, activation='softmax'))` "
   , what are meanings of the values of the arguments ?

   (ii) What will be the dimension of the feature map produced after applying the second convolution layer?

   d) Consider the Keras code given above in question no. 8(d) and answer the following questions.
   In the line containing "model.add(Dense(500, activation='relu'))", if activation='sigmoid' is used, will this CNN network work? If it works, explain what will be the difference from the previous case, but if it does not work, explain why?   5 x 4 = 20 marks

----