The LNM Institute of Information Technology

Computer Science and Engineering

Deep Learning (CSE4121)

Mid Term

Time: 90 minutes | Date: September 26, 2019 | Max. Marks: 50

Read the following instructions carefully:

- *There are 5 questions printed on both sides of the paper.*
- **No marks** *for providing just expressions/answers unless accompanied with correct justification and/or derivation.*
- *In case of any doubt, make your assumption, write it clearly and continue.*

1. (a) Write loss function (i.e., for one example) for logistic regression with each and every variable explained.
   (b) Evaluate the value of loss for the above function, in the cases when:
      i. Prediction is 1 and label is 0.
      ii. Prediction is 0 and label is 0.
   (c) Assume that there are only two classes in your data. The data is such that no linear function can separate them reasonably. Discuss the behaviour of logistic regression for such datasets.

   3+2+2+2 Marks

2. (a) Write gradient descent (ONLY algorithm) for a neural network that is L layers deep. Explain all the variables clearly.
   (b) If two hidden layers in the above neural network do not use any activation and only compute the linear function of the output of previous layer. What impact it may have on the depth of the network? Justify.
   (c) What is the need for epochs in gradient descent algorithm?

   4+(1+2)+3 Marks

3. (a) In a convolutional neural network (CNN), if the filter size is $f \times f$, the size of the image is $n \times n$, padding used is p and a stride of s is also applied. Derive a formula for the dimensions of the output.

   (b) Let the input volume to a convolutional layer have dimensions $6 \times 6 \times 3$. You want to have convolution with a filter that has size $3 \times 3 \times 2$. Suggest a way to be able to do so. Also find the dimensions of the output if no padding is used and stride is 1.

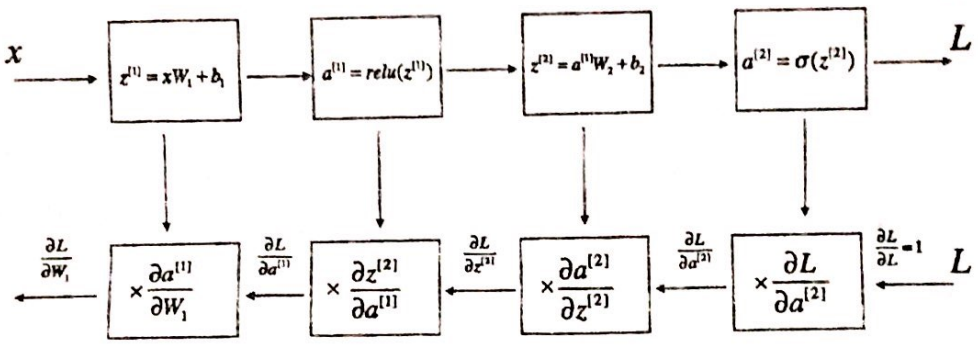
5+(3+2) Marks

4. A 2-layer neural network is provided as described next. Size of the input layer is 3, first hidden layer is 2 and output layer is 1. The hidden layer is using ReLU and output layer is using sigmoid as activation.

   (a) Draw a neat and clean neural network for the above description.

   (b) Provide fully labeled and detailed weight and bias matrices for the network so drawn.

   (c) Use the following figure to provide $dW^{[1]}$, $db^{[1]}$, $dW^{[2]}$ and $db^{[2]}$. You need to show calculations only for one parameter of each matrix. Note that $dW^{[1]}$ is a short form for $\frac{\partial L}{\partial W^{[1]}}$, where L is the loss.



2+4*1+4*1 Marks

5. Consider the following model of a neural network as defined in TensorFlow using Keras module.

```
model = tf.keras.models.Sequential([
  tf.keras.layers.Conv2D(10, (3,3), activation='relu', input_shape=(34, 34, 1)),
  tf.keras.layers.MaxPooling2D(2, 2),
  tf.keras.layers.Flatten(),
  tf.keras.layers.Dense(100, activation='relu'),
  tf.keras.layers.Dense(1, activation='sigmoid')
])
```

   (a) Draw the neural network corresponding to this code.

   (b) Assuming there are 5 layers in the above model (one corresponding to each `tf.keras.layers` statement), what is the number of parameters that can be learnt at each layer?

   (c) How does the answer to 5(b) change if `MaxPooling2D(4,4)` is used instead of `MaxPooling2D(2,2)`?

4+5*1+2 Marks