The LNM Institute of Information Technology
Computer Science and Engineering
Deep Learning (DL)
Mid Term

Time: 90 minutes   Date: September 28, 2018   Max. Marks: 50

Read the following instructions carefully:

- *There are 6 questions printed on both sides of the paper.*

- *Answer to each part of a question that carries 2 marks must be limited to 50 words. If you exceed the limit, you may get lesser marks.*

- **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. Why is deep learning taking off right now? Draw the figure that we had seen in the class on *Performance* vs *Amount of data*. In the context of the figure, elaborate your answers based on the following:

    (a) Computational resources

    (b) Existence of data

    (c) Existence of the Internet

    (d) Advances in the field itself

    2+2+2+2 Marks

2. Why do we need activation functions in neural networks? Derive the mathematical functions for these activation functions using their properties.

    (a) Sigmoid

    (b) Tanh

    (c) Leaky relu (Also comment on the differentiability of the function)

    2+(2+2+2) Marks

3. Write and justify the cost function for logistic regression gradient descent. What do we mean by learning in the context of gradient descent, i.e., when can learning be said to be over? Also write the equations that update parameters for gradient descent.

    Hint: Cost function is different from the squared error distance and to justify it, consider how it penalizes any wrong prediction.

    (2+2)+2+2 Marks

4. Discuss bias and variance problems for a deep neural network. Which of these problems can be solved using regularization? Explain dropout.
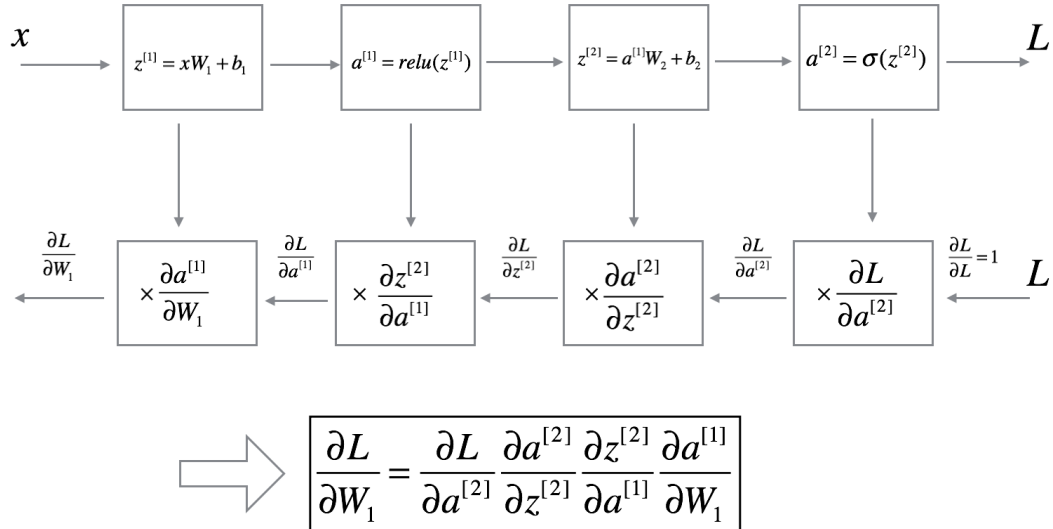
    (2+2)+2+2 Marks

5. State the intuition behind the backpropagation algorithm for neural networks. How does it help in learning? What are the parameters that it helps find? How is chain rule for differentiation helpful in backpropagation?

   Hint: A figure is attached to help you answer. Ignore any numbers written in the figure.

$$\frac{\partial L}{\partial W_1} = \frac{\partial L}{\partial a^{[2]}} \frac{\partial a^{[2]}}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial a^{[1]}} \frac{\partial a^{[1]}}{\partial W_1}$$

2+2+2+2 Marks

6. Consider a neural network as shown in the figure.
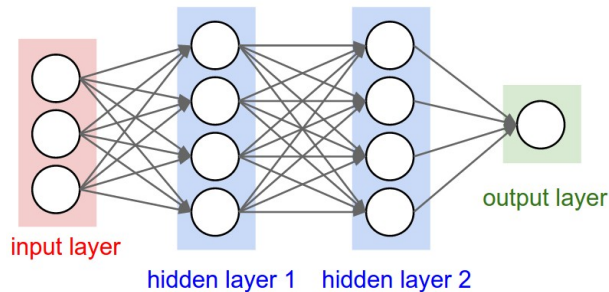
   (a) How many parameters need to be learned in the above neural network?

   (b) How should the parameters be initialized for the learning algorithm?

   (c) Label the figure neatly with z and a for each node of all the layers.

   (d) Write the vectorized form for the answers computed in part (c). Use actual matrices, just capitalizing is not enough.

   (e) Is it correct to use tanh as activation function in hidden layer 1, relu in hidden layer 2 and sigmoid in the output layer?

2+2+2+2+2 Marks