Masters in Science

MCSE-304: Deep Learning

Semester III

Time: 45 minutes **Max. Marks: 40**

1. Derive a gradient descent training rule for a two-layer network with a single output unit employing sigmoid activation function. The network is supplied with m training instances. The intermediate layer employs ReLu activation function.
   The network uses the cost function for logistic regression given by:
   $$E(t,y) = -(y\log(t) + (1-y)\log(1-t))$$
   where t is the target output and y is the neural network output. 10

2. The tanh activation usually works better than sigmoid activation function for hidden units because the mean of its output is closer to zero, and so it centers the data better for the next layer. True/False? Justify your answer. 2

3. You are building a binary classifier for recognizing cucumbers (y=1) vs. watermelons (y=0). Which one activation functions would you recommend using for the output layer? ReLU, Leaky ReLU, sigmoid, tanh? Justify your answer. 2

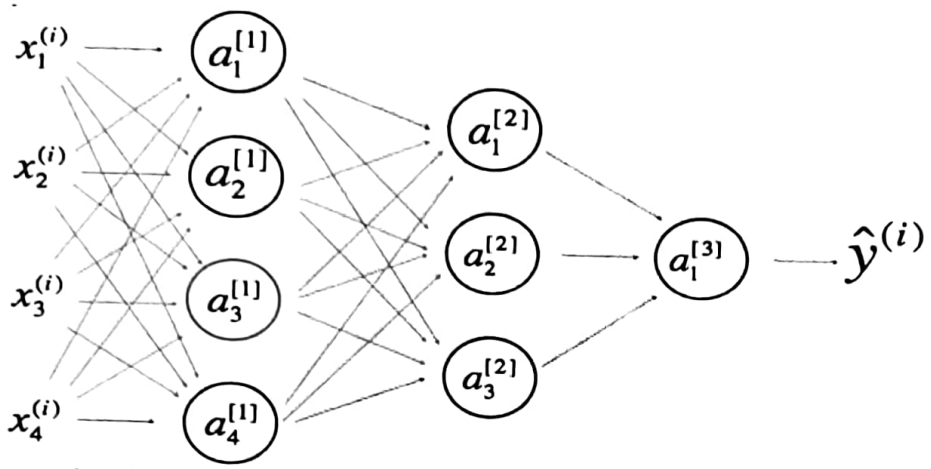
4. Consider the following code:
   ```
   A = np.random.randn(4,3)
   B = np.sum(A, axis = 1, keepdims = True)
   ```
   What will be B.shape? 2

5. Suppose you have built a neural network. You decide to initialize the weights and biases to be zero. Which of the following statements are True? (Check all that apply). 4

- Each neuron in the first hidden layer will perform the same computation. So even after multiple iterations of gradient descent each neuron in the layer will be computing the same thing as other neurons.
- Each neuron in the first hidden layer will perform the same computation in the first iteration. But after one iteration of gradient descent they will learn to compute different things because we have "broken symmetry".
- Each neuron in the first hidden layer will compute the same thing, but neurons in different layers will compute different things, thus we have accomplished "symmetry breaking" as described in lecture.
- The first hidden layer's neurons will perform different computations from each other even in the first iteration; their parameters will thus keep evolving in their own way.

6. What do you mean by term "hyperparameters" in a neural network? List any four. What happens if the learning rate is too high or too low? 4

7. Consider the following neural network:



Determine determine the dimensions of following arrays: 6

a) W[1]
b) b[1]
c) W[2]
d) b[2]
e) b[3]
f) W[3]

5. Perform average pooling on the following input using a 3×3 filter and stride 2:

9	25	13	1	18
24	12	5	18	6
11	4	17	11	22
2	15	8	20	14
18	7	20	13	22

The following Convolutional Neural Network (CNN) is applied to a database of images to predict the true class of the image out of 10 choices. Each image is of size 100 × 100 with 3 channels. Give the dimensions of the input layer of the CNN. If all convolution layers have stride 1 and all pooling functions have stride 2, determine the total number of parameters for each layer.

Input layer: As mentioned in the question above.
C1: Conv 3×3, padding same, 10 filters
C2: Conv 7×7, padding same, 20 filters
P1: 2×2 max pooling
C3: Conv 5×5, padding same, 10 filters
P2: 4×4 max pooling
FC1: fully connected
Output layer: Softmax

Note: C1, C2, C3 denote convolution layers, P1 and P2 denote pooling layers, and FC1 denotes fully connected layer. 10