CIS 419/519: Quiz 8

November 7, 2019

1. Consider two neural networks, A and B, which are both a series of k fully connected layers. Each layer is the same size in A and B. The only difference is that A has a non-linear function in between each layer and B has no non-linear activations at all.

   What can you say about the size of each network's hypothesis space?

   (a) A's is larger than B's

   (b) A's is the same as B's

   (c) A's is smaller than B's

   (d) Not enough information

   Answer: a

2. Assume we have a black-and-white image that is 16x16x1 (16 pixels wide, 16 pixels high, and 1 color channel) and a convolutional neural network that is 7x7, has 1 input channel, 5 output channels, stride 1, and padding 0.

   How many parameters does the CNN have? Do not count any bias terms.

   (a) 256

   (b) 245

   (c) 12544

   (d) 500

   Answer: b

3. Assume you are training a multi-layer neural network that has some non-linear activation function in between each layer. Do the values you use to initialize the parameters change the final parameter values you find after training?

   (a) Yes. Because the optimization is non-convex, it could find a different local optima based on the starting location.

(b) There is not enough information to decide.

(c) No. Given an appropriate learning rate for gradient descent, the optimization will always converge to the same solution.

(d) Yes. The network described above is optimizing a convex objective function. Given an appropriate learning rate, gradient descent will always find the global optimum.

Answer: a

4. Assume we have a black-and-white image that is 16x16x1 (16 pixels wide, 16 pixels high, and 1 color channel) and a fully connected layer that has output size 10x10x1.

   How many parameters does the fully connected layer have? Do not count any bias parameters.

   (a) 160

   (b) 100

   (c) 25600

   (d) 1600

   Answer: c

5. Say we are training a 1 layer neural net with 1 output unit. The net input to a unit is defined as $net_j = \sum w_{ij} \cdot x_i$ where net_j is the one final output unit with inputs from several input units. The output is defined as a sigmoid output $o_j = \frac{1}{1+e^{-net_j - T_j}}$ where T_j is our threshold value. The measured error is $\frac{1}{2}(t_j - o_j)^2$ where t_j is our true label.

   What is the update rule to some weight w_{ij}?

   (a) $\Delta w_{ij} = R(\frac{1}{2}(t_j - o_j)^2)$

   (b) $\Delta w_{ij} = R(\frac{1}{2}(t_j - o_j)^2)(\frac{1}{1+e^{-net_j - T_j}})(\sum w_{ij} \cdot x_i)$

   (c) $\Delta w_{ij} = R(t_j - o_j)o_j(1 - o_j)x_i$

   (d) $\Delta w_{ij} = R(t_j - o_j)x_i$

   Answer: c