**Miniproject 2**
**BMI 555 IEE 520**
**Fall 2020**

**Due Date: September 29, 2020**

1) Consider a neural network to classify the *Avila* data. Complete the following calculations **without** a neural network software package, so you understand the steps. You might use *Microsoft Excel*.

Consider a network with an input layer, 3 hidden layers with 50, 40, 30 nodes in the hidden layers, and an output layer with one node for each class and a softmax function applied to the output.

a) How many parameters are estimated in this network (including constants)?

b) Consider the following instance of data:

| x1 | x2 | x3 | x4 | x5 |
|---|---|---|---|---|
| 0.266074 | -0.16562 | 0.32098 | 0.483299 | 0.17234 |
| x6 | x7 | x8 | x9 | x10 |
| 0.273364 | 0.371178 | 0.929823 | 0.251173 | 0.159345 |

Suppose the weights from the input layer to the first node of the first hidden layer are:

| w1 | w2 | w3 | w4 | w5 | |
|---|---|---|---|---|---|
| 0.040065 | 0.246356 | 0.234578 | 0.172821 | 0.449315 | |
| w6 | w7 | w8 | w9 | w10 | b |
| 0.875052 | 0.321868 | 0.141179 | 0.128957 | 0.264436 | 0.031945 |

And a sigmoid activation function is applied.

Calculate the value at this node after the sigmoid function is applied.

c) Suppose that for an instance, the output *before* a softmax function is applied is

| A | B | C | D | E | F |
|---|---|---|---|---|---|
| 1.197033 | 1.79776 | 1.952738 | 1.23094 | 1.065328 | 0.41491 |
| G | H | I | W | X | Y |
| 1.212204 | 1.86248 | 1.74077 | 1.484271 | 1.950335 | 1.657855 |

Calculate the output *after* a softmax function is applied and estimate the class probabilities for this instance.

d) For the instance in part (c), what class is assigned?

e) For the instance in part (c), suppose that the actual target value for this instance is Class *A*. Calculate the contribution to the cross entropy loss function from this instance.

2) Consider a neural network to classify images of 128 x 128 handwritten digits in one of 10 classes with a *softmax* function at the output. For each network below, calculate how many parameters are estimated (including constants).

a) A fully connected network is used with 3 hidden layers and 50, 40, 30 nodes in these layers.

b) A network with local connectivity (but not weight sharing), 3 hidden layers, 1 8x8 filter between input and hidden layers, stride = 4, (with a different constant weight for each hidden node), and a fully connected output layer.

c) A network with local connectivity, and weight sharing (including a shared constant weight) (convolutional network), 1 filter 8x8 between input and hidden layers, stride = 4, and a fully connected output layer.

d) A network with local connectivity, and weight sharing (including the constant weight) (convolutional network) for each of 8 filters between input and all hidden nodes, all strides = 4, with A,B,C,D filters using 8x8 patches and E,F,G,H filters using 4x4 patches in each layer, and a fully connected output layer.

3) Use the *Avila* dataset. Use the neural network package in **SKLEARN ONLY**, to design the best network for me to use on similar data in the future. Also, I want to know about the quality of your model. Provide an estimate the generalization error of the model you recommend.

Submit

- Your final code, your output, and your estimate of generalization error
- A brief description of what models you explored and how you estimated generalization error. You are writing for me, so you can summarize based on my knowledge of the topic.
- A clear statement of architecture and **any parameters changed from the SKLEARN defaults**, so I could build your network. It is important to learn to communicate your models clearly.