Student Name: ______________________________

Student Number: ______________________________

Answer the questions in the spaces provided on the question sheets. If you run out of room for an answer, continue on the back of the page.

Question	Points	Score
1	10	
2	20	
3	10	
4	40	
5	5	
6	15	
7	10	
8	10	
Total:	120	

Graduate students must answer the GR question. The EC question is for extra credit. Undergraduate students may answer the GR or the EC question for extra-credit. Total possible points are 110/100 for undergraduate and 120/110 for graduate students.

[10 Pts] 1. **True/False.** Answer the following true/false questions.

(a) ____ Regularization helps the network avoid overfitting on training data.

(b) ____ The formula for softmax loss function is $\frac{e^{-s_k}}{\sum_i e^{-s_i}}$

(c) ____ In stochastic gradient descent the loss is calculated over all training data.

(d) ____ 1×1 convolutional kernels allow the data to be interpolated across channels from the previous layer.

(e) ____ The receptive field of 2 successive 3×3 convolutional kernels is equal to that of one 5×5 kernel.

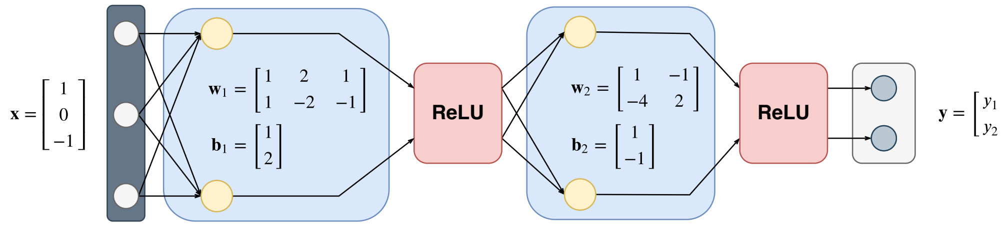
[20 Pts] 2. What is the decision boundary implemented by a neural network with N inputs, 1 output, an no hidden layers? Explain with equation(s).

[10 Pts] 3. Prove the following about activation function derivatives.

(a) $\sigma'(x) = \sigma(x)(1 - \sigma(x))$

(b) $\tanh'(x) = 1 - \tanh^2(x)$

4. A 2-layer perceptron is designed to classify objects with 3-dimensional features into two classes.



[5 Pts] (a) Write the equation of this perceptron only in terms of matrix multiplications and ReLU functions. What are the parameters of each matrix multiplications $\hat{\mathbf{x}}, \hat{\mathbf{w}}_1, \hat{\mathbf{w}}_2$ in terms of $\mathbf{x}$, $\mathbf{w_1}$, $\mathbf{w_2}$, $\mathbf{b_1}$, and $\mathbf{b_2}$?

[5 Pts] (b) Draw the computational graph of this neural network. Hint: Use only matrix multiplications. Use Concatenation nodes. Use ReLU nodes.

[2 Pts] (c) On the computational graph write down the equation for each node and label each edge with its dimension.

[8 Pts] (d) Perform the forward pass on this graph for the input given in the figure –i.e., $\mathbf{x} = [1, 0, -1]^T$.

[20 Pts] (e) Perform the backward pass of the backpropagation on this graph. Hint: The derivative of ReLU function for an input x_i is 1 if $x_i \geq 0$, otherwise it is 0. For example, if $\mathbf{x} = [-1, 3, 2, -3]^T$ then $\text{ReLU}'(\mathbf{x}) = [0, 1, 1, 0]^T$.

[5 Pts] 5. Explain three advantages of using convolutional kernels in deep neural networks over fully connected layers.

[15 Pts] 6. Consider a CNN whose inputs are RGB images of size 512×512. The network has two convolutional layers. Answer the following questions about this network:

(a) If the first layer has 12 convolutional filters, the spatial dimensions of the first layer feature maps are 504×504 and no padding is applied, what are the sizes of the first layer's kernels?

(a) ______________

(b) If a max-pooling sizes of 2×2 and stride 2 is applied in the first layer, what are the sizes of the first layer's pooled feature maps?

(b) ______________

(c) What is the depth of the pooled feature maps?

(c) ______________

(d) If the second layer contains 6 kernels of size 3×3 and no padding is applied, what is the size of the second layer feature maps?

(d) ______________

(e) If the second layer also contains a 2×2 with stride 2 pooling layer, what are the final dimensions of the vectors that result from vectorization (unwrapping) of the last layer of the CNN?

(e) ______________

[10 **GR**] 7. Specify the structure, weights, and bias(es) of the smallest neural network capable of performing the minimum distance classification on two tightly grouped and linearly separable classes with means m_1 and m_2 in N-dimensions.

Hint: The decision boundary equation for minimum distance classification is:

$$d_{12}(\mathbf{x}) = d_1(\mathbf{x}) - d_2(\mathbf{x}) \tag{1}$$

where:

$$d_i(\mathbf{x}) = \mathbf{m}_i^T \mathbf{x} - \frac{1}{2}\mathbf{m}_i^T \mathbf{m}_i \tag{2}$$

[10 **EC**] 8. Suppose that the input to a convolutional kernel is an image with dimensions $N \times N$. The size of the kernel is $F \times F$, and stride is S, with padding of P pixels applied on both sides of the image.

(a) Prove the size of the output image is $M \times M$, where:

$$M = \frac{N + 2P - F}{S} + 1 \tag{3}$$

(b) Prove that the amount of padding to keep the output image size the same as the input image size is:

$$P = \frac{(S-1)N + F - S}{2} \tag{4}$$