ECE/CS 559 - Fall 2018 - Midterm #1.

Full Name: **ID Number:**

Q1 (30 pts). Let u be the step activation function with $u(x) = 1$ if $x \geq 0$, and $u(x) = 0$, otherwise. Consider the perceptron $y = u(w_0 + w_1 x_1 + w_2 x_2)$, where w_1 and w_2 are the weights for inputs x_1 and x_2, respectively, w_0 is the perceptron bias, and y is the perceptron output. Let $\mathcal{C}_0 = \{\begin{bmatrix}0 & 0\end{bmatrix}, \begin{bmatrix}0 & 1\end{bmatrix}, \begin{bmatrix}1 & 0\end{bmatrix}\}$, and $\mathcal{C}_1 = \{\begin{bmatrix}1 & 1\end{bmatrix}\}$. The desired output for class $\mathcal{C}_0$ is 0, and the desired output for class $\mathcal{C}_1$ is 1. Correspondingly, let $d(\mathbf{x}) = 0$ if $\mathbf{x} \in \mathcal{C}_0$, and otherwise, let $d(\mathbf{x}) = 1$ if $\mathbf{x} \in \mathcal{C}_1$.

(a) **(10 pts)** If possible, find w_0, w_1, w_2 that can separate $\mathcal{C}_0$ and $\mathcal{C}_1$ (i.e., provide the desired output for all 4 possible input vectors). Otherwise, prove that no choice of weights can separate the two classes.

(b) **(10 pts)** Recall that the perceptron training algorithm relies on the update $\mathbf{w} \leftarrow \mathbf{w} + \eta(d(\mathbf{x}) - y)\begin{bmatrix}1 & \mathbf{x}\end{bmatrix}$, where $\mathbf{w} = \begin{bmatrix}w_0 & w_1 & w_2\end{bmatrix}$ is the weight vector. Let $\eta = 1$ and the initial weight vector be given by $\mathbf{w} = \begin{bmatrix}-0.5 & 1 & 0\end{bmatrix}$. Calculate the updated weights after one epoch of training.

.

(c) **(10 pts)** Will the weights provided by the algorithm (as setup in (b)) eventually converge after a sufficiently larger number of epochs? Justify your answer.

Q2 (40 pts): Consider the activation function

$$f(v) = \begin{cases} v, & v \geq 0, \\ 0, & v < 0. \end{cases}$$

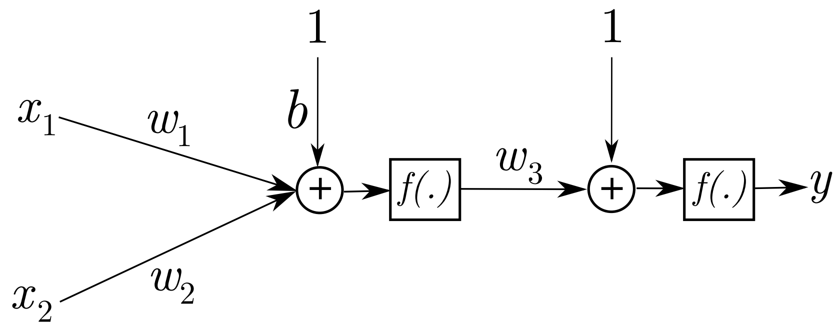
This is also known as the rectified linear unit (ReLU). For (a)-(c), consider a single-neuron network with input-output relationship $y = f(b + \mathbf{w}^T \mathbf{x})$, where y is the network output, b is the bias term, $\mathbf{w} = \left[\begin{smallmatrix} w_1 \\ w_2 \end{smallmatrix}\right]$ are the synaptic weights, and $\mathbf{x} = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]$ is the network input.

(a) **(5 pts):** Calculate the derivative $f'(v)$ for every $v \neq 0$.

(b) **(10 pts):** Let $E = (d - y)^2$, where d is a constant (a generic desired output). Write down the delta-learning rule (the gradient-descent update equations) for b, w_1, w_2 given learning parameter $\eta = \frac{1}{2}$.

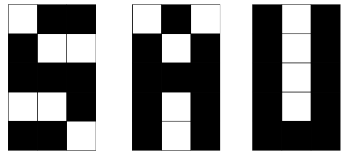
(c) **(10 pts):** Consider the same delta-learning setup as in (b). Consider the training vectors $\mathbf{x}_1 = \left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right]$, $\mathbf{x}_2 = \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$, with desired outputs $d_1 = -1$, $d_2 = 2$, respectively. Find the updated bias and the updated weights after 2018 epochs of online learning given initial conditions $b = 0$, $\mathbf{w} = \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$.

(d) **(15 pts)** Consider now a multi-layer network as shown below.

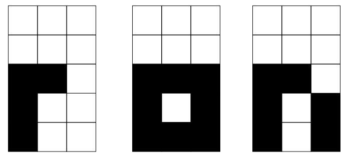


Let $E = e^{(d-y)^4}$. Find the gradient-descent update equations for b, w_1, w_2, w_3 given $\eta = \frac{1}{4}$.

Q3 (30 pts). [Hint: This question does not require any calculations or long answers. If you do find yourself doing calculations or coming up with long answers, you are in the wrong path.] In a classification problem, Class $\mathcal{C}_0$ contains the following three 5×3 black and white images:



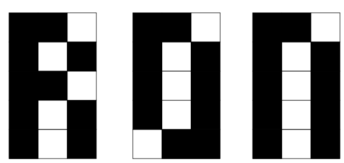
On the other hand, class $\mathcal{C}_1$ contains the following images:



A black pixel corresponds to a "1" and a white pixel corresponds to a "0." Equivalently, both $\mathcal{C}_0$ and $\mathcal{C}_1$ are sets of three 15-dimensional vectors.

(a) **(10 pts)** Recall that the step activation function u is given by $u(x) = 1$, $x \geq 0$, and $u(x) = 0$, otherwise. Design a neural network (i.e. find an appropriate network topology together with its weights and biases) whose neurons use the step activation function u such that the network provides an output of 0 for members of class $\mathcal{C}_0$, and it provides an output of 1 for members of class $\mathcal{C}_1$. In your answer, label the 15 inputs of the network as $x_{11}, \ldots, x_{53}$, where $x_{ij} \in \{0, 1\}$ represents the pixel value at the ith row jth column of the image. For example, the "r" member of class $\mathcal{C}_0$ would be encoded as $x_{31} = x_{41} = x_{51} = x_{32} = 1$, and $x_{11} = x_{12} = x_{13} = x_{21} = x_{22} = x_{23} = x_{33} = x_{42} = x_{43} = x_{52} = x_{53} = 0$.

(b) **(10 pts)** Repeat (a) for the case where the class $\mathcal{C}_1$ is given as follows:



(c) **(10 pts)** Consider the class $\mathcal{C}_1$ as shown in (b) that consists of the capital letters "R," "O," and "N." Design a single-layer network with 15 inputs (a 5×3 binary image as before) and 3 outputs such that

(1) the network output (from the three neurons) is given by $[1 \;\; 0 \;\; 0]^T$ if the input pattern is "R,"

(2) the network output is given by $[0 \;\; 1 \;\; 0]^T$ if the input pattern is "O," and

(3) the network output is given by $[0 \;\; 0 \;\; 1]^T$ if the input pattern is "N."

Hint: All three images in (b) have exactly 10 black pixels.