CS/ECE/ME532 Activity 24

1. **Neural net functions**

   a) Sketch the function generated by the following 3-neuron ReLU neural network.

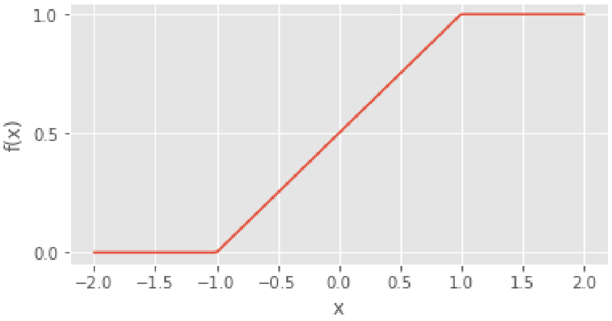
   $$f(x) = 2(x - 0.5)_+ - 2(2x - 1)_+ + 4(0.5x - 2)+$$

   where $x \in \mathbb{R}$ and where $(z)_+ = max(0, z)$ for any $z \in \mathbb{R}$. Note that this is a single-input, single-output function. Plot $f(x)$ vs x by hand.

   b) Consider the continuous function depicted below. Approximate this function with ReLU neural network with 2 neurons. The function should be in the form

   $$f(x) = \sum_{j=1}^{2} v_j (w_j x + b_j)_+$$

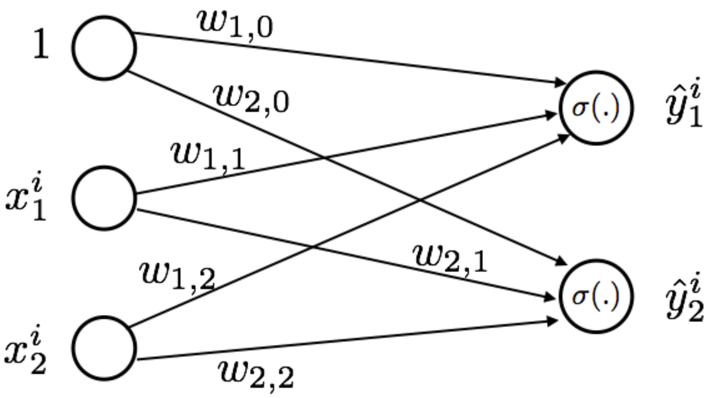
   Indicate the weights and biases of each neuron and draw the corresponding neural network diagram.



2. A script is available to train two neurons using stochastic gradient descent to solve two different classification problems. The two classifier structures are shown below. Here we use a logistic activation function $\sigma(z) = (1 + e^{-z})^{-1}$. The code generates training data and labels corresponding to two decision boundaries: $x_2^i = -2x_1^i + 0.2$, and $x_2^i = 5(x_1^i)^3$.

   a) Do you expect that a single neuron will be able to accurately classify data from case 1? Why or why not? Explain the impact of the bias term associated with $w_{1,0}$.

   b) Do you expect that a single neuron will be able to accurately classify data from case 2? Why or why not? Explain the impact of the bias term associated with $w_{2,0}$.

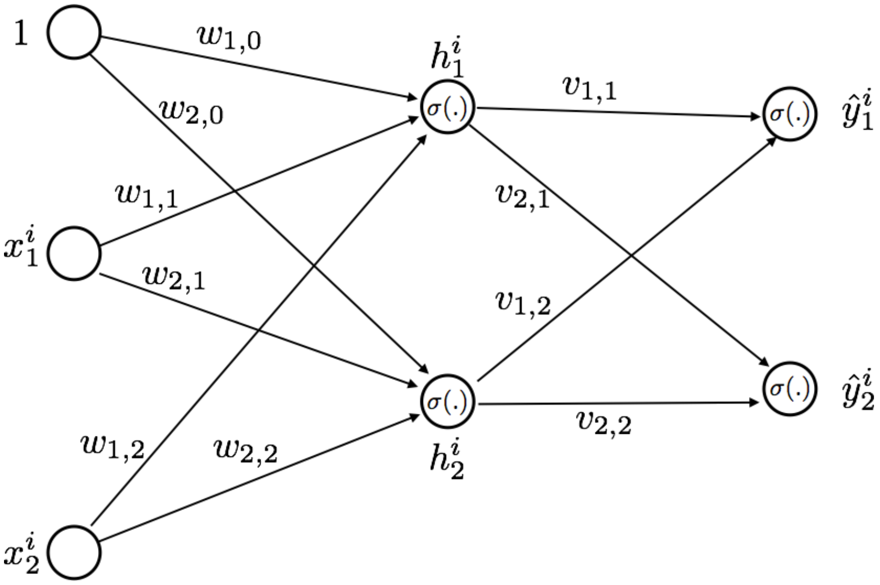


c) Run SGD for one epoch. This means you cycle through all the training data one time, in random order. Repeat this five times and find the average number of errors in cases 1 and 2.

d) Run SGD over twenty epochs. This means you cycle through all the training data twenty times, in random order. Repeat this five times and find the average number of errors in cases 1 and 2.

e) Explain the differences in classification performance for the two cases that result with both one and twenty epochs.

3. This remainder of this activity uses a three-layer neural network with three input nodes and two output nodes to solve two classification problems. We will vary the number of hidden nodes. The figure below depicts the structure when there are two hidden nodes.



A second script is available that generates training data and trains the network using SGD assuming a logistic activation function $\sigma(z) = (1 + e^{-z})^{-1}$.

a) Use $M = 2$ hidden nodes and ten epochs in SGD. Run this four or five times and comment on the performance of the two classifiers and whether it varies from run to run.

b) Repeat $M = 2$ but use 100 epochs in SGD. (You may use fewer epochs if it takes more than a minute or two per run.) Run this several times and comment on the performance of the classifiers and whether it varies from run to run.

c) Recall the two-layer network results from the previous problem. How do the possible decision boundaries change when you add a hidden layer?

d) Now use $M = 3$ hidden nodes and run 100 epochs of SGD (or as many as you can compute). Does going from two to three hidden nodes affect classifier performance?

e) Repeat the previous part for $M = 4$ hidden nodes and comment on classifier performance.