Homework 04 B - 50 points

Due December 5, 2019

Provide a PDF file entitled hw04.pdf that answers all the questions below and shows the figures generated by your code. For parts 1 and 2, please show your work (typed submission is preferred). Submit the PDF via Canvas and include it in your Github repository.

1 Output Layer

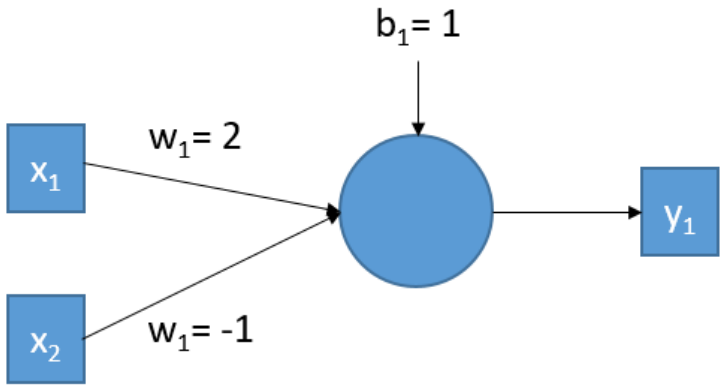

Figure 1: Network with Only Output Layer

Given a single training sample $X = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, desired label $d = 1$ and the above model:

1. Calculate the output, y_1, of the network using a sigmoid activation function.
2. Using the chain rule, derive the gradient for w_i when the loss function is squared error, $J = \frac{1}{2}\sum_{i=1}^{N}(d_i - y_i)^2$.
3. Compute the updates for w_1, w_2, and b_1 with $\eta = 1$ using stochastic gradient descent.

2 Single Hidden Layer

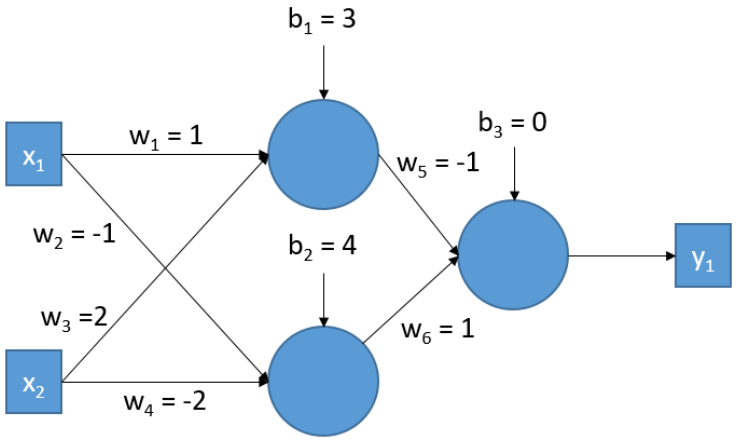

Figure 2: Network with One Hidden Layer

Given a single training sample $X = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, desired label $d = 0$ and the above model:

1. Calculate the output, y_1, of the network using a sigmoid activation function.
2. Using the chain rule, derive the gradient for w_j (j^{th} weight on the input to the hidden layer) when the loss function is squared error, $J = \frac{1}{2}\sum_{i=1}^{N}(d_i - y_i)^2$.
3. Compute the updates for w_1, w_2, and w_5 with $\eta = 1$ using error back-propagation with stochastic gradient descent.

3 UF Network

Figure 3: Training Data for UF Network

For this section of the assignment, you will need to create a 2-hidden layer network to distinguish the letters and the background. Starter code is provided for you but please make sure you have Pytorch installed. Make sure your code is well commented and include your name at the top of the script. You will need to determine the number of units for each hidden layer as well two other hyperparamters: the learning rate and number of epochs. Answer the following questions:

1. How many units will you need in the first and second hidden layers? Why?
2. To determine the hyperparameters, please select three different learning rates and number of epochs and report the values selected. Also, generate the a) learning curve and b) decision boundary for the top three combinations of the different learning rates and number of epochs (total of 6 figures).
3. For the best learning rate and number epochs from the above plots, provide discussion on the a) influence of each hyperparameter on training performance and b) why the selected values worked well.