HW #6 Due: 5/24/2023

1. We mentioned in the lecture that the perceptron (pp. 8 of back propagation) with a loss function of $J(\boldsymbol{w}) = |z_k d_k| - z_k d_k$ can be trained with the following updating rule:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \begin{cases} 0, & \text{if } z_k d_k > 0 \\ \eta \boldsymbol{x}_k d_k, & \text{otherwise} \end{cases}$$

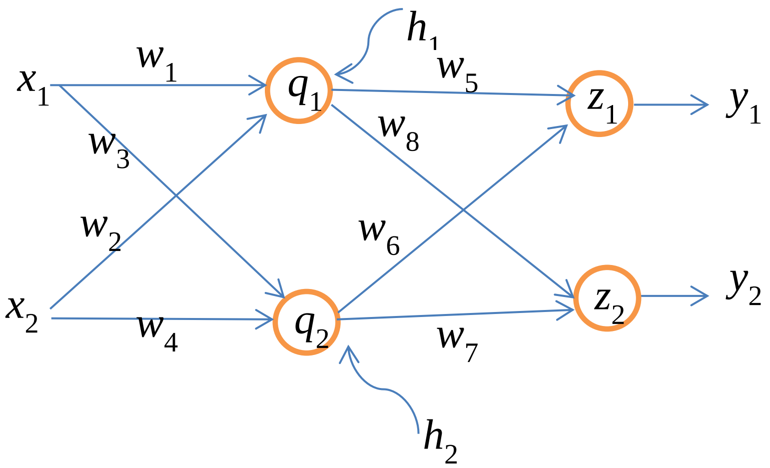
Show that this algorithm is directly derived from the stochastic gradient descent algorithm.

2. For the neural network given below, let $w_1 = 2.0$, w_2 to w_8 be 1.0, $d_1 = 1.0$, $d_2 = 0.0, \eta = 0.1, x_1 = 1.0,$ and $x_2 = -1.0$. The activation function from q_1 to h_1 and q_2 to h_2 is ReLU, the activation function at the output nodes is linear (i.e., $y = z$), and the cost function is

$$J = \frac{1}{2}\sum_{i=1}^{2}(y_i - d_i)^2.$$

(i) Find y_1 and y_2 (forward computation).

(ii) Find the value of $\Delta w_1 = \eta \frac{\partial J}{\partial w_1}$ by using the BP algorithm.



3. Prove the following statement: The derivative of the softmax activation function has the following form

$$\frac{\partial}{\partial z_i} y_\ell = \begin{cases} y_\ell(1 - y_\ell), & \text{if } i = \ell \\ -y_\ell y_i, & \text{if } i \neq \ell \end{cases}$$

4. Reproduce the MSE plot on pp. 9 of the 23_Cost functions. The input is 1.0, initial weight $w = 2.0$, bias $b = 2.0$, $\eta = 0.15$, and desired output is 0.0. Because this problem is simple, you can manually compute the updating equations for $w(k+1)$ and $b(k+1)$, and write a hard-coded program without using a library, such as Keras.

5. Implement a perceptron with the learning rule given in Problem 1 to classify the classes of virginica and Versicolor in the Iris dataset. Use a bias term in the perceptron to make it a linear binary classifier. As usual, take 70% of the samples as the training set and then report the average accuracy after 10 trials. You need to determine a suitable value of η and the number of training epochs.