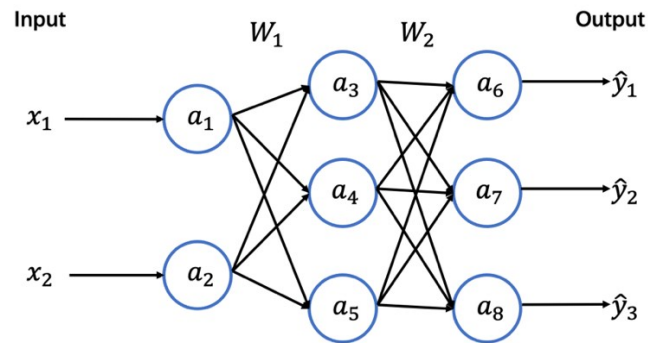


1. In this question, we consider a multi-class classification problem, i.e., to discriminate C ($C \geq 3$) classes. We use softmax function in the output layer to produce probability predictions for each class. In specific, consider the following neural network with 2 input neurons, 3 neurons in hidden layer and 3 output neurons. For each hidden neuron, we use *ReLU* activation function. The input and output of this neural network is represented by $\mathbf{x} \in \mathbb{R}^2$ and $\hat{\mathbf{y}} \in \mathbb{R}^3$. The weight matrix from the input layer to hidden layer is denoted by $W_1 \in \mathbb{R}^{2 \times 3}$ and the weight matrix from the hidden layer to output layer is denoted by $W_2 \in \mathbb{R}^{3 \times 3}$.

Here, we use one-hot encoding to convert ground truth labels into “one-hot” vectors, where only an entry is 1 and other entries are 0. For example, suppose there are C classes in total, the one-hot encoding of a label c (an integer) is a n -dimensional vector where only the c -th entry is equal to 1 while other entries are 0. We use loss function of Negative Log Likelihood (NLL) to optimize the network. Given the ground truth label in the one-hot encoding format $\mathbf{y} = [y_1, y_2, y_3]$ and the probability prediction of the network $\mathbf{p} = \text{softmax}([\hat{y}_1, \hat{y}_2, \hat{y}_3]) = [p_1, p_2, p_3]$, the NLL loss between the ground truth label and probability prediction is defined as $\ell = -\sum_{i=1}^3 y_i \log p_i$.



- (a) Given a training set and test set listed in *train_q4.csv* and *test_q4.csv*, write PyTorch code to learn the above network, and present the followings: 1) plot the curve of training loss, 2) give the final optimized weights for (W_1, W_2) , 3) test the network on the given test set and report the testing accuracy.
- (b) The learning rate is an important hyperparameter when training a neural network. Let's study this hyperparameter here. Please re-train the network with learning rates in the set $\{1.0, 0.1, 0.01\}$, respectively. Meanwhile, observe the training loss curve and testing accuracy for each learning rate.

Questions: 1) plot the training loss curve and report the testing accuracy for each learning rate; 2) what is your finding from these training loss curves and testing accuracy?