I. What is the problem?

In this homework, we have to build a 3-layer neural network to recognize handwritten digits from images, which is likely to be more complex, more precise than the fewer-layer neural network we've built in previous homework.

III. Please note that the image stored in the dataset is already reshaped into a vector. Now, please play with the code in A2 to understand the data and answer the following questions.

a. What is the dimensions (or shape or size) of TRimg. Please explain the meaning of the values in each dimension.

   TRimg has the shape of (60000, 784). The array has 60000 elements. Each element is an array of size 784, which is the vectorized value of the corresponding image in the training dataset

b. What is the dimensions (or shape or size) of TSlab. Please explain the meaning of the values in each dimension.

   TSlab has the shape of (10000,). The array has 60000 elements. Each element is a value from 0-9 that is the correct label of the corresponding image in the testing dataset

V. Please use Google/Chatgpt/Ch3.2 to answer the following:

a. What is training dataset and test dataset.

   - The training dataset is the set of data points with input features and corresponding correct labels for each point. This dataset will be used to train the model to learn the underlying pattern and relationships
   - The testing dataset, similar to the training dataset, also contains data points and labels, however this dataset is not used in the training process. It will be used to test the performance of the trained model.
   - These datasets are separated to prevent the model from memorizing the training data instead of learning the patterns.

b. How do we use the two datasets?

   Mentioned above

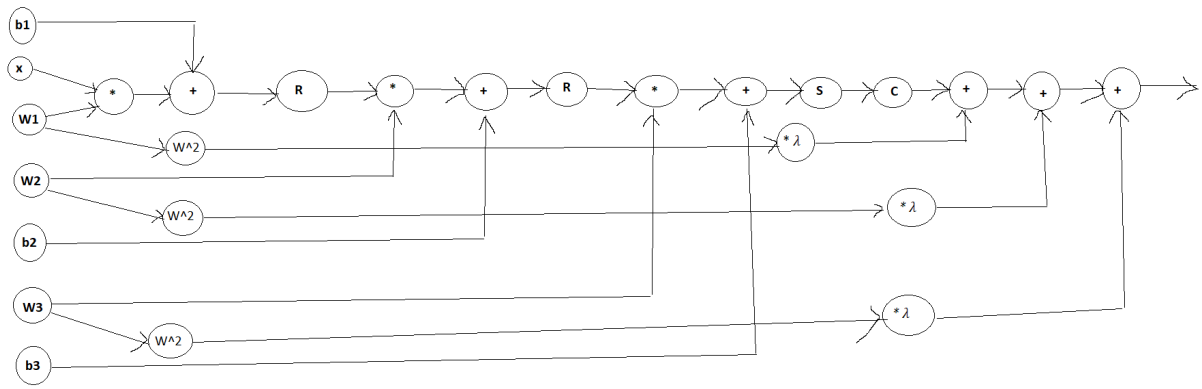
c. What is batch and what is epoch in machine learning?
   - Typically, the entire dataset is too large to process at once, so it is divided into smaller subsets, called batches. The model will update the weights after processing each batch.
   - Epoch is a complete pass through the entire training dataset. During an epoch, the model will go through every step, from processing data, updating weights, calculating loss, and evaluating performance.

- Both are hyperparameters that have to be tuned manually.

VI. What is the model we want to use in this homework? Please write down the equations following the examples in the “Chapter 4.4 Neural Network Demos V2.pdf” (Page 43, 47,48)?

softmax(ReLU(y = ReLU(y = x * W1 + b1) * W2 + b2) * W3 + b3)

VII. Please follow the examples in the “Chapter 4.4 Neural Network Demos V2.pdf” (Page 17-37,& 50-51)_ to draw the computational graph and write down the forward and backward propagation (computation for each node).



VIII. What are the parameters used in this homework (parameter not hyperparameter!!)? And how do you plan to define the parameters in the code.

Following the model in question VI, the parameters will be W1, b1, W2, b2, W3, b3. The parameters will be initialized as such:

```
W1=0.01*np.random.randn(D, Hid1)
b1=np.zeros((1,Hid1))
W2=0.01*np.random.randn(Hid1,Hid2)
b2=np.zeros((1,Hid2))
W3=0.01*np.random.randn(Hid2,K)
b3=np.zeros((1,K))
```

Where Hid1 is the number of nodes in the first hidden layer, Hid2 is the number of nodes in the second hidden layer, K is the number of labels (10 in this homework), D is the number of features in the input layers (784 in this homework)