The learning objectives for Project 3 are:

1. Make use of word2vec embeddings to classify the same tweets as in Project 2 into 3 classes (positive, negative and neutral).
2. Make use of cross validation
3. Make use of PyTorch DataSet
4. Make use of PyTorch DataLoader
5. Make use of layers from the nn module of PyTorch (nn.embeddings, nn.linear, nn, softmax, nn.relu)
6. Learn to write a forward function
7. Make use of loss functions from the nn module of PyTorch
8. Understand auto differentiation and make use of it
9. Make use of Adam optimizer
10. Generate confusion matrix and repeat the analysis from Project 2.

The links below provide introductory materials for completing the above tasks.

<https://nadbordrozd.github.io/blog/2016/05/20/text-classification-with-word2vec/>

<https://medium.com/@martinpella/how-to-use-pre-trained-word-embeddings-in-pytorch-71ca59249f76>

<https://pytorch.org/tutorials/beginner/basics/data_tutorial.html>

<https://stackoverflow.com/questions/69967363/scikit-learn-train-test-split-into-pytorch-dataloader>

<https://discuss.pytorch.org/t/cross-validation-in-pytorch/120182> Read the code in the first post. They are doing validation for every training epoch. You can do every 50 epochs or so.

I have included code for constructing a multi-layer neural network for a simple (non-NLP) problem using the components of the nn module mentioned above in Pytorch-Simple.ipynb. The associated data is in data.txt. You may want to figure out how to add a DataLoader to this simpler setting. Then you can follow almost the same logic to build your NLP model after incorporating ideas from the other links provided to build your model.

The first link above describes end-to-end model building. But, doesn’t use PyTorch.

The second link uses PyTorch and shows how to include an embedding layer. But uses GRU and doesn’t use linear and softmax.

The third link shows how to use DataSet and DataLoader. But, does it for images.

The Pytorch-Simple.ipynb shows everything I want. But, is not set up for NLP and doesn’t use softmax.

However, taken together all of them provide what you need to do.