DMML Assignment 2

Name: Sampad Kumar Kar
Shankar Ram Vasudevan

Summary

- We use semi-supervised learning to learn a fairly accurate neural network based on the fashion MNIST dataset.
- We employed the following strategies of semi-supervised learning:
    - **Strategy 1:** We train our model on just 50 randomly labelled instances. We obtain an accuracy of 52%.
    - **Strategy 2:** We use clustering strategy to create newer labels for learning. This enables us to label the majority of unlabelled data with just a few labels. We did that in the following ways:
        - **Strategy 2.1:** We first cluster the training set into 50 clusters and for each cluster we select the image closest to the centroid point and choose it as the representative image of the cluster. Then we collect these 50 images and manually label them and train the model on these. This gives us an accuracy of 65%.
        - **Strategy 2.2**: We go one step further from **Strategy 2.1** and propagate the labels of the representative example to all the points in its cluster and then use the whole labelled set as the training data. This gives an accuracy of 67%.
        - **Strategy 2.3:** Instead of propagating the label of representative to all of the points in the cluster as in **Strategy 2.2**, we instead propagate the labels to the 75th percentile closest to the centroid. This gives us a huge improvement and takes the accuracy to 88%.
- We also employ supervised learning and use the original training set to train the model. We achieve a meagre improvement of 1% over **Strategy 2.3**. This demonstrates the advantage and powers of semi-supervised learning in the machine learning paradigm.