Fashion MNIST challenge

Team member: Solo (Swagat Bhowmik)

I am using Gradient boosting model to predict labels for the test images. I am using gradient boosting since it can be used for classification problems and the problem in hand required classifying the images into 10 categories. Also Gradient boosting is quite flexible with regards to optimization and can handle noise in the dataset well.

Preparing the dataset:

First I am loading the train and test data from their .npy files as NumPy arrays. And then normalizing it so that each pixel value is between 0 and 255.

I load the labels as well from the training and testing label csv files.

Then I reshape the NumPy array to be of 2 dimensions so that it can be fed into the model.

Challenge:

It was tricky to come up with a model to use for the problem because of several available options for classification problems.

I was thinking of using XGBoost to implement gradient boosting which is like an optimized implementation of it. It has built in hyperparameter tuning and also, better at minimizing overfitting. But for some reason the labels I generated were continuous decimal values instead of discrete integers in between 0 to 9.

Also because of the large size of the dataset it was taking a significant amount of time to train the model. Hence, I couldn't explore many options.

Learning outcome:

I used Gradient Boosting for first time and also got familiar with Kaggle.

Accuracy:

Got a public score of 0.82 with generated labels.

Work to be done:

I also plan to make XGBoost work properly and also use a CNN model to see the difference in results.