ME 532 Final Project: Update 1

This first update includes initial data visualization and analyses. The fashion MNIST data was already split as into training and testing datasets. The following table gives a brief outlook of the data set.

	label	pixel1	pixel2	pixel3	pixel4	pixel5	pixel6	pixel7	pixel8	pixel9	...	pixel775	pixel776	pixel777	pixel778	pixel779	pixel780	pixel781	pixel782	pix
0	2	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	
1	9	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	
2	6	0	0	0	0	0	0	0	5	0	...	0	0	0	30	43	0	0	0	
3	0	0	0	0	1	2	0	0	0	0	...	3	0	0	0	0	1	0	0	
4	3	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	

5 rows × 785 columns

As summarized in the problem statement the dataset has been classified into 10 labels corresponding to different clothing items as shown:

The approach towards has been slightly changed to what was previously planned. The k-means and PCA algorithms will be implemented first and then regression methods will be tried.

The change was made after looking into the dataset with bit more detail. K-means clustering seemed the first algorithm to start with and will probably take the least time.
The scripting of the algorithm has been started and the first step in implementation of k-means would be identify the number of clusters. Since the number of items are 10 in the problem set. It seemed a logical starting point to choose k=10.

The initial progress has been slow since it took time in scripting the k-means algorithm.
In the coming updates the results of k-means algorithm will be shared, and this will be followed by regression and neural networks.