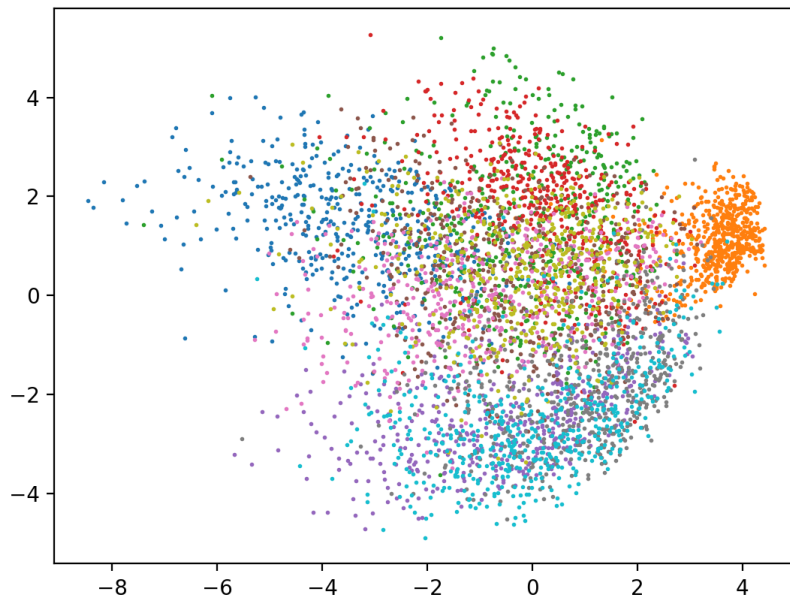


## Homework 5 – Machine Learning (CS453X, Whitehill, Spring 2019)

You may complete this homework assignment either individually or in teams up to 2 people. You should use the same MNIST dataset as you did for Homework #3.

1. **Principal Component Analysis (PCA) [30 points]:** In this problem you will implement Principal Component Analysis (PCA). Note that you are required to implement PCA from scratch; you may not use any off-the-shelf software (e.g., `sklearn`). Apply the PCA algorithm you implemented to the 10,000 examples in the MNIST *test set*. (There's no deep conceptual reason for this choice; it's just because the test set is smaller and will reduce the computation time.) In particular, for each image  $\mathbf{x}$  in the MNIST test set, project it onto the *first* and *second* principal components, i.e., the two eigenvectors of  $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^\top$  with the highest and second-highest associated eigenvalues. Call these two projections  $p_1$  and  $p_2$ ; then plot  $(p_1, p_2)$  of each MNIST image on a 2-D plane, where the *color* of each point is determined by the class label of that image  $\mathbf{x}$ . You should obtain an image similar to the following. Note, however, that the colors in your image could differ (since they are arbitrary); also, the *orientation* of the image could also be different (since the eigenvectors of a given matrix are not unique).



You should attach the figure that you rendered as either a PNG or a PDF.

Submit your Python code in a file called `homework5_WPIUSERNAME1.py` (or `homework5_WPIUSERNAME1_WPIUSERNAME2.py` for teams). Submit your figure (either a PDF or PNG) using an analogous naming convention.