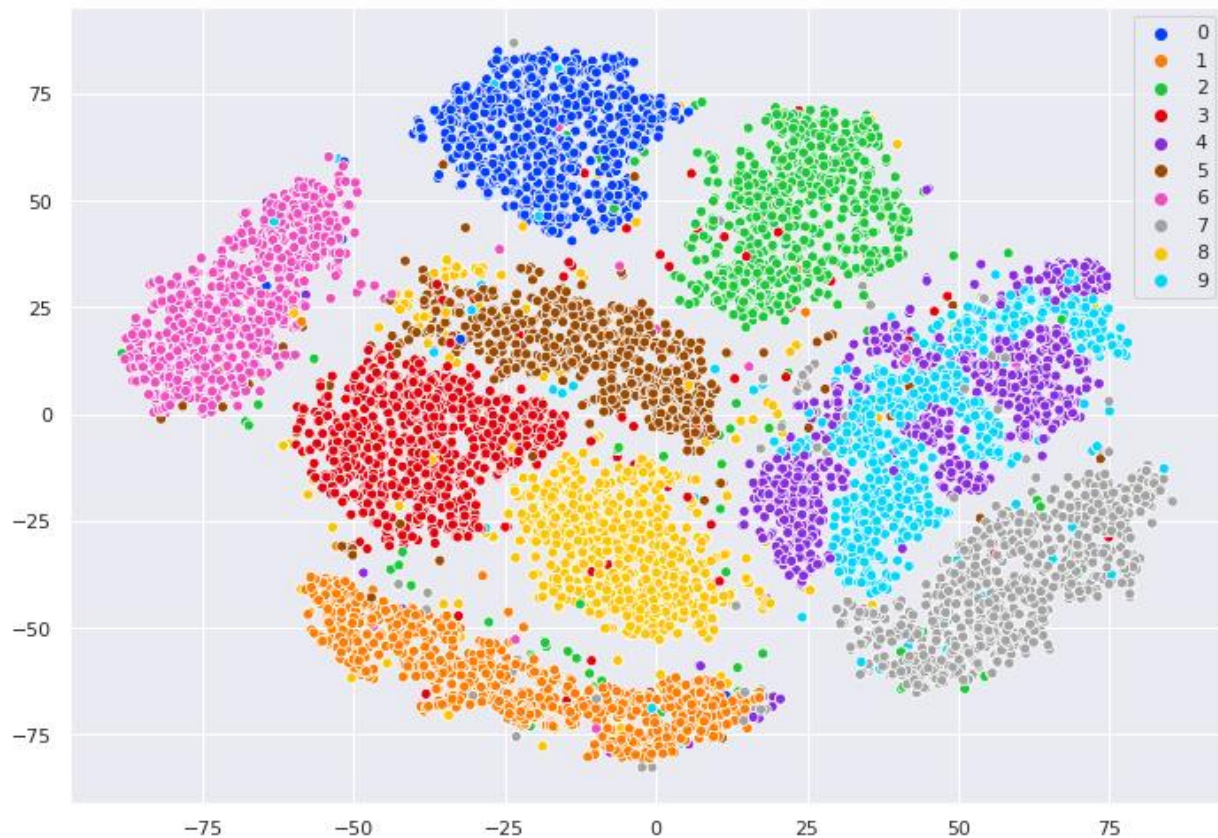


Assignment 2

Fuzzy Clustering and Cluster Scoring.



In this assignment you will use the same imbalanced data set taken from MNIST as you did in Assignment #1. The t-SNE'd version of a random different version of similar data is shown above.

For this assignment you will

- 1) implement your own version of the Fuzzy C-Means clustering algorithm, "from scratch".
- 2) you will apply this to both the dataset in the file (with 784 dimensions), as well as a PCA'd version of the data where you've reduced the dimensionality down to 2.
- 3) you will then write your own version of the Adjusted Rand Index algorithm (from scratch), and use it to score the results of your clustering on both versions of the data.

NOTE: you do not need to write PCA from scratch.

You will turn in both the code and a short report, discussing what you've implemented, how well it worked, what you've learned, etc. Remember, you must understand what you turn in – you may be asked to explain it to me and/or the class.

NOTE: Do *not* just comment your code and then turn that in as your report. I will deduct points. Your report will usually have *at least* a couple of pages of text, figures, etc.