Ream me file

Run all those files: just do python [file\_name]

Please run this commend

1. **python -m SimpleHTTPServer 8888**: please run this command to open up the **mnist.html in** [**http://localhost:8888/mnist.html**](http://localhost:8888/mnist.html)

**Note: please use Chrome as browser (did not test in other browsers)**

This project is to compare PCA and TSNE dimension reduction methods.

**Download\_mnist.py** is used to download the data from online

**Tsne\_mnist.py** is used to generate coordinates of all data points from high dimension to 2-dimension plane using TSNE method

**Pca\_mnist.py** is used to generate coordinates of all data points from high dimension to 2-dimension plane using PCA method

**Vis\_mnis.py** is used to generate the pictures of all data points and store them to PNG files in pictures folder.

1. For **extra credits** and **Extensions**:

I created a file called **functionalityExtension.html.**

You can use the **three buttons** for three visualizations.

**PCA, TSNE, and the Cluster using Forces**

The group one uses the **forces around a circle**, and the label image would be displayed in the middle.

**Midata** is the raw data scraped from the internet

**Picture** is the folder containing all pictures generated from vis\_mnis file

**PCA\_coords.csv**: file stores results from **Pca\_mnist.py**

**TSNE\_coords.csv:** file stores results from **Tsne\_mnist.py**

Mnist.html, the final html file that include all visualization for both **pca** and **tsne**.

**Detailed Summary:**

**PCA and TSNE are methods of dimensionality reductions.**

**Comparing the TSNE and PCA, those two methods use different strategies. TSNE applies the covariance matrices and statistical approximation to scatter points and project points into lower dimension from Gaussian Distribution to small-samples T distribution. PCA uses the main component vectors. The first vector is the highest variance vector, and the second one is the second highest variance vector. Thus, points are related somehow through two directions along those two vectors. (in x and y in our visualizations)**