

Homework# 7 - Due day 11:59PM, 17th January, 2018

- t-SNE

- ▶ There are nice implementations of t-SNE in different programming languages: <https://lvdmaaten.github.io/tsne/> the implementation contains a trick to compute squared euclidean distance for all pairs of input, ref: <https://stackoverflow.com/questions/37009647>

25 points ▶ Try to modify the code a little bit and make it back to symmetric SNE (therefore you need to first understand how the t-SNE is implemented and find out the specific code piece to modify)

25 points ▶ Try to visualize the embedding of both t-SNE and SNE and discuss their differences. (please use the data included in https://lvdmaaten.github.io/tsne/code/tsne_python.zip, there is a file named “mnist2500_X.txt” which contains 2500 feature vectors with length 784, for describing 2500 mnist images. Another file named “mnist2500_labels.txt” provides corresponding labels) crowding problem might not happen

50 bonus ▶ Try to visualize the distribution of pairwise similarities in both high-dimensional space and low-dimensional space, based on both t-SNE and symmetric SNE

25 points ▶ Try to play with different settings of perplexity, and see if there is any change in visualization.

75 points ▶ submit a report: what you have done (code, testing performance), what you have visualized, what you have learned.