ASSIGNMENT 6

CS5304 - T-SNE

1. ASSIGNMENT

In this assignment, you'll explore data visualizations using t-SNE. The Dataset you'll use is Fashion MNIST by Zalando. This is about 60,000 fashion images in grayscale spanning 10 different classes. This dataset in built into PyTorch torchvision datasets. You can also find it on GitHub if you want it in Numpy format (https://github.com/zalandoresearch/fashion-mnist).

You'll implement the following 4 visualizations and in your write-up discuss your findings, the relative merits and disadvantages of each approach. All of these visualizations should be done in 2D and we'll expect to see one colorized image per projection. See the example notebook we post in the Slack channel to see how we colorize the image. Also, to save some compute time, you can subsample the data set to a smaller size when you are doing t-SNE embeddings.

- PCA projection
- ISOMAP
- t-SNE on the raw image
- t-SNE on features extracted from either resnet18 or FashionSimpleNet trained on Fashion MNIST

For the last task above, here is an implementation of these models that can be trained on Fashion MNIST (https://github.com/kefth/fashion-mnist). Train one of these models on Fashion MNIST and then use it as a feature extractor by extracting the embeddings from the penultimate layer (the input layer to the final FC layer). Use these embeddings to perform t-SNE.

In your analysis, you should at least answer the following questions:

- How is a linear projection (PCA) different than the non-linear projection (ISOMAP)?
- How does t-SNE differ from PCA?
- How does dimensionality reduction on the raw data compare to the same process on the extracted features?

2. GRADING RUBRIC

- Proper implementation of all 4 visualizations - 6 points (1.5 points per visualization)
- Writeup and analysis - 6 points
- Bug-free code - 3 points

3. Regrading Policy

We'll accept regrade requests for 1 week after releasing the assignment grade. Please email Andrew (apd64@cornell.edu) with why you think your assignment should be regraded. We're also happy to clarify the meaning of comments. We will regrade the entire assignment, and it is possible that after the regrade your assignment might have a lower score.