t-Distribution Stochastic Neighborhood Embedding - t-SNE

Introduction -

PCA is good for dimensionality reduction and visualization for handling multi-dimension dataset. **But, what if you could use something more advanced than PCA?** We will know about a new algorithm called t-SNE (2008), which is much more effective than PCA (1933).

Table of Content -

1. What is t-SNE?
2. What is dimensionality reduction?
3. How does t-SNE fit in the dimensionality reduction algorithm space?
4. How t-SNE Works ?
5. Use cases -
6. Where and When to use t-SNE?
7. Comparison PCA vs T-SNE

1. What is t-SNE?

t-SNE is a non-linear dimensionality reduction algorithm used for exploring high-dimensional data. It maps multi-dimensional data to two or more dimensions suitable for human observation.

2. What is dimensionality reduction?

It is the technique of representing multi-dimensional data (data with multiple features having a correlation with each other) in 2 or 3 dimensions.

3. How does t-SNE fit in the dimensionality reduction algorithm space?

Limitations of PCA -

PCA is a linear algorithm. It will not be able to interpret complex polynomial relationship between features. On the other hand, t-SNE is based on probability distributions with random walk on neighborhood graphs to find the structure within the data.

A major problem with, Linear dimensionality reduction algorithms like PCA does it focuses on placing dissimilar data points far from each other. **But does not place similar points close to each other.**

Local approaches seek to map nearby points to nearby points in the low-dimensional representation. Global approaches preserve geometry at all scales, i.e mapping nearby points to nearby points and far away points to far away points.

t-SNE is capable of retaining both the Local and Global structure at the same time.
Most of the Non-Linear techniques are not capable of retaining both structure.

4. How t-SNE Works ?

It finds patterns in the data by identifying observed clusters based on similarity of data points with multiple features.

t-SNE do following things-

1. Expands dense cluster
2. Shrink sparse cluster
3. Does not preserve distance betw cluster
4. Try with multiple value of perplexity
5. Run iteration till shape stabilize

Perplexity means no. of neighbor. Result of t-sne may vary even in same input. In t-sne, S stands for Stochastic which means Probabilistic whose result are not always fixed.

5. Use cases -

t-SNE can be used on almost all high dimensional data sets. But it is extensively applied in Image processing, NLP, genomic data and speech processing. It has been utilized for improving the analysis of brain and heart scans. Below are a few examples-

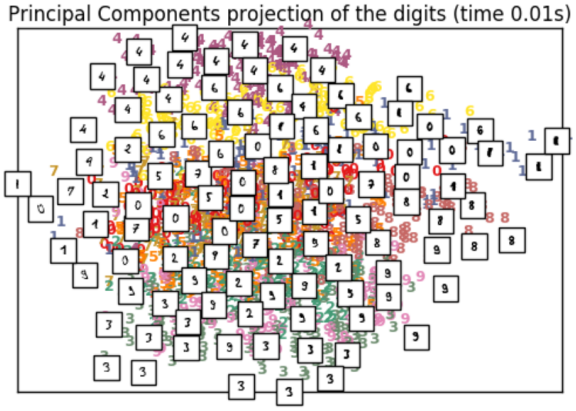
1. Facial Expression Recognition
2. Identifying Tumor subpopulations (Medical Imaging)
3. Text comparison using wordvec

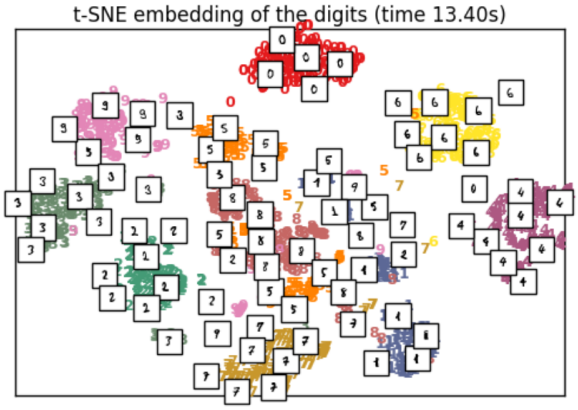
6. Where and When to use t-SNE?

Problem with the t-SNE algorithm is that it doesn’t always provide a similar output on successive runs.

So then how could you use the algorithm? The best way to used the algorithm is to use it for exploratory data analysis. It will give you a very good sense of patterns hidden inside the data. It can also be used as an input parameter for other classification & clustering algorithms.

Comparsion PCA vs TSNE


Principal Components projection of the digits (time 0.01s)


t-SNE embedding of the digits (time 13.40s)