1 t-SNE

t-SNE which stands for t distribution-Stochastic neighborhood embedding. Suppose you have 50 dimensional data set, as it is like impossible task for us to visualize and get a sense of it. We have to convert that 50 dimensional data set to something which we can visualize or with which we can play around. This is where t-SNE comes into the picture it converts the higher dimensional data into the lower dimensional data by following steps:

- It measures the similarity between the two data points and it does for every pair. Similar data points will have more value of similarity and the different data points will have less value.
- Then it converts that similarity distance to probability(joint probability) according to the normal distribution.
- It does the similarity check for every point. Thus it will have the similarity matrix 'S1' for every point. This is all calculation it does for our data points that lie in higher dimensional space.
- Now, t-SNE arranges all of the data points randomly on the required lower dimensional (let's suppose 2).
- And it does all of the same calculation for lower dimensional data points as it does for higher ones - calculating similarity distance but with a major difference it assigns probability according to t- distribution instead of normal distribution and this is because it is called t-SNE not simple SNE.
- Now we also have the similarity matrix for lower dimensional data points. Let's call it S2.
- Now what t-SNE does is it compares matrix S1 and S2 and tries to make the difference in between matrix S1 and S2 much more smaller by doing some complex mathematics.
- At the end we will have lower dimensional data points which tries to capture even complex relationships at which PCA fails.
- So on a very high level this is how t-SNE works.

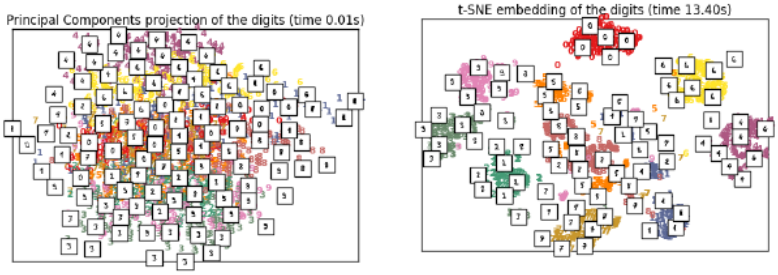

Figure 1

References:

For more details on t-SNE,
`https://en.wikipedia.org/wiki/T-distributed_stochastic_neighbor_embedding`
For t-distribution,
`http://mathworld.wolfram.com/Studentst-Distribution.html`
For Normal Distribution,
`https://en.wikipedia.org/wiki/Normal_distribution`