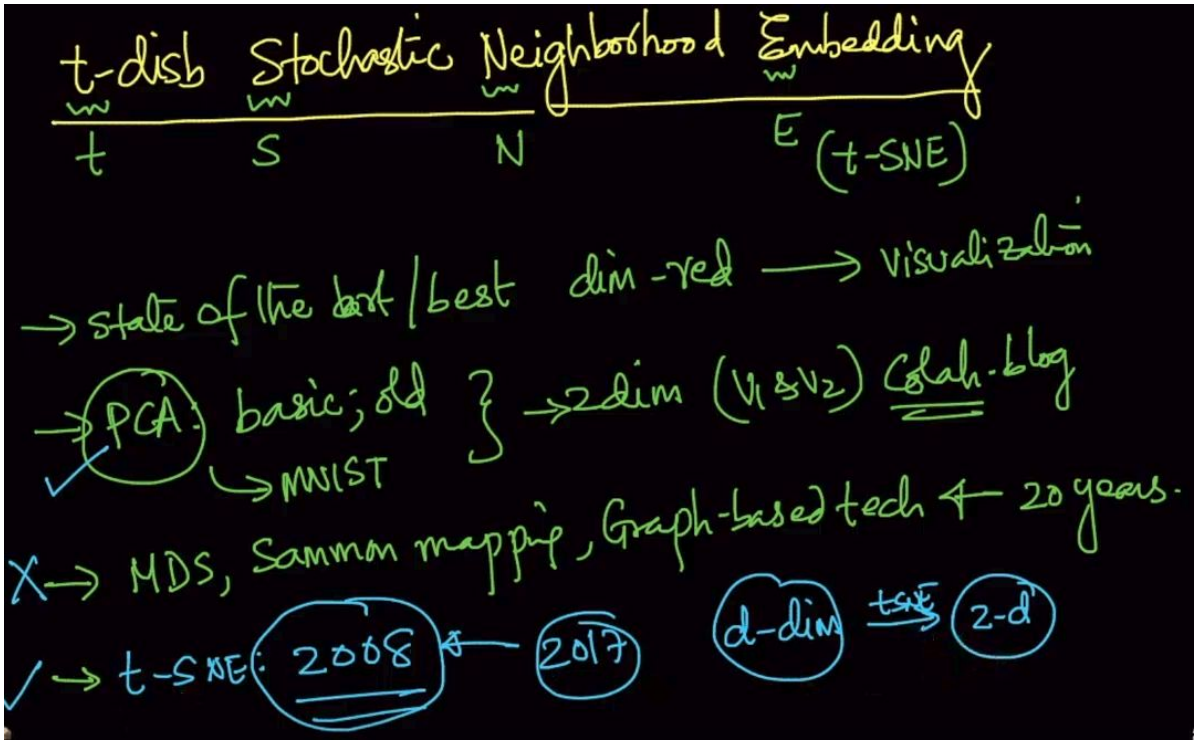
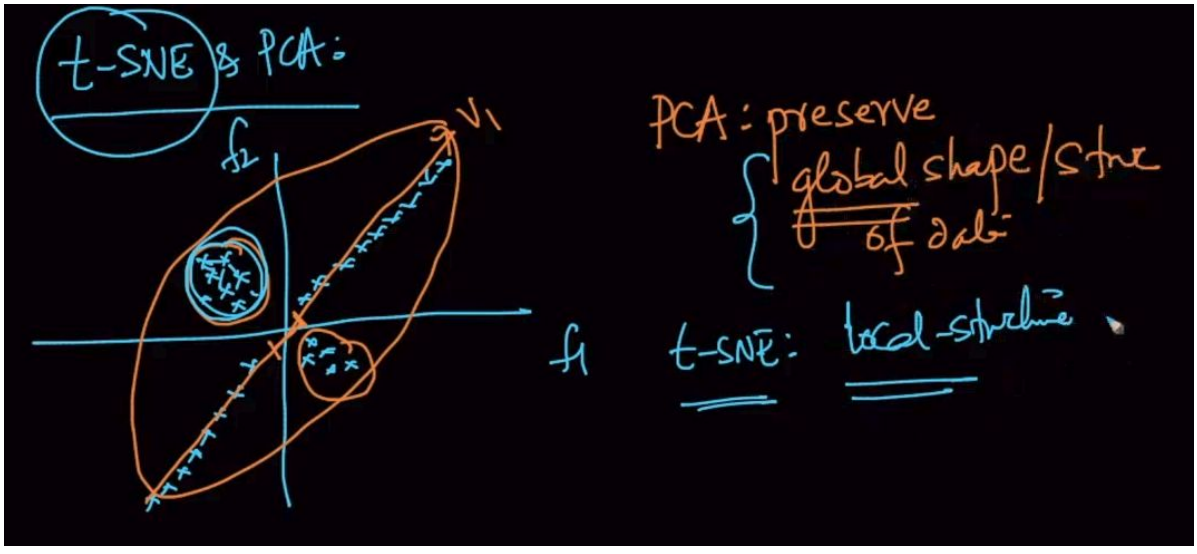
t-SNE(T-distributed Stochastic Neighborhood Embedding)



t-SNE is state of the art /best dimensionality reduction technique for visualization.

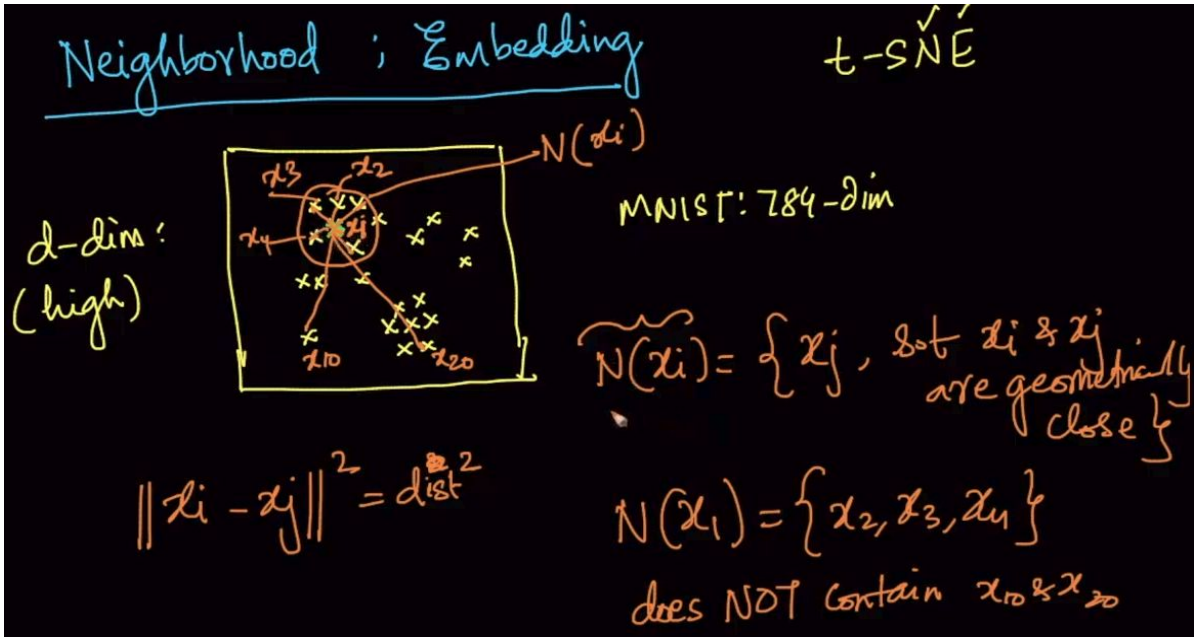
DIFFERENCE BETWEEN PCA AND t-SNE



PCA : Preserves global shape/structure of data

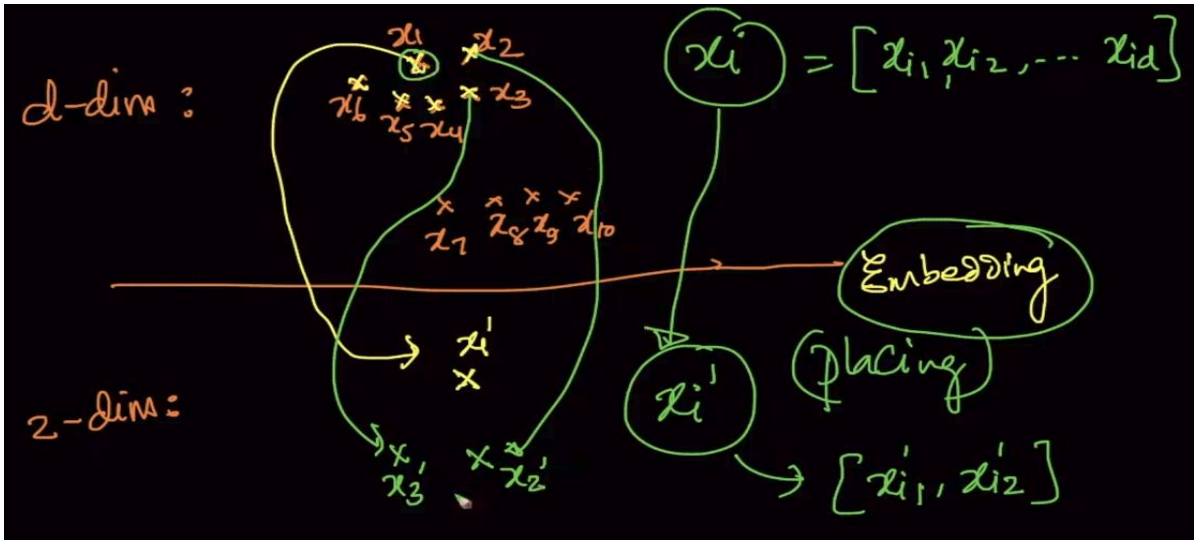
t-SNE : Local Structure

NEIGHBORHOOD



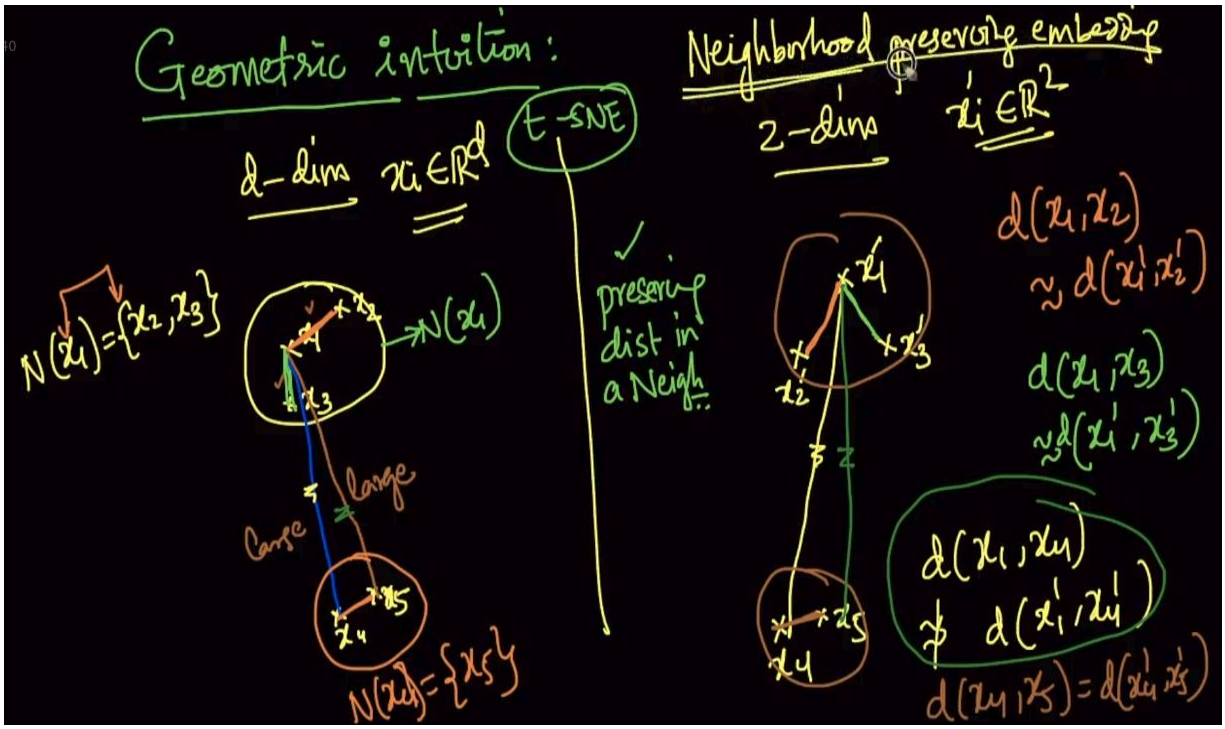
Neighborhood (N) is all the points that are closer to the data point/ Distance between datapoints is small. Here, x_i a data point and it's N = $\{x_2, x_3, x_4\}$

EMBEDDING



Embedding means taking a datapoint in a high dimensional space and placing it in a low dimensional space/ finding corresponding point in low dimensional space.

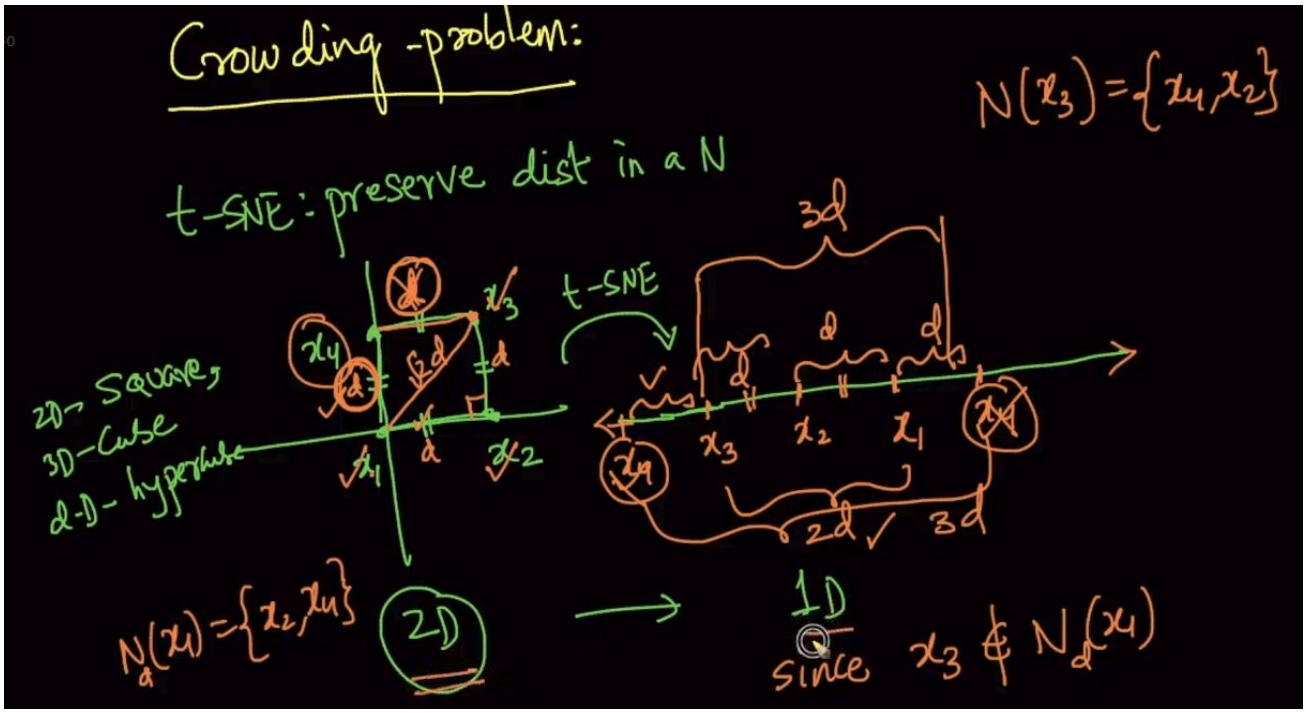
GEOMETRIC INTUITION of t-SNE



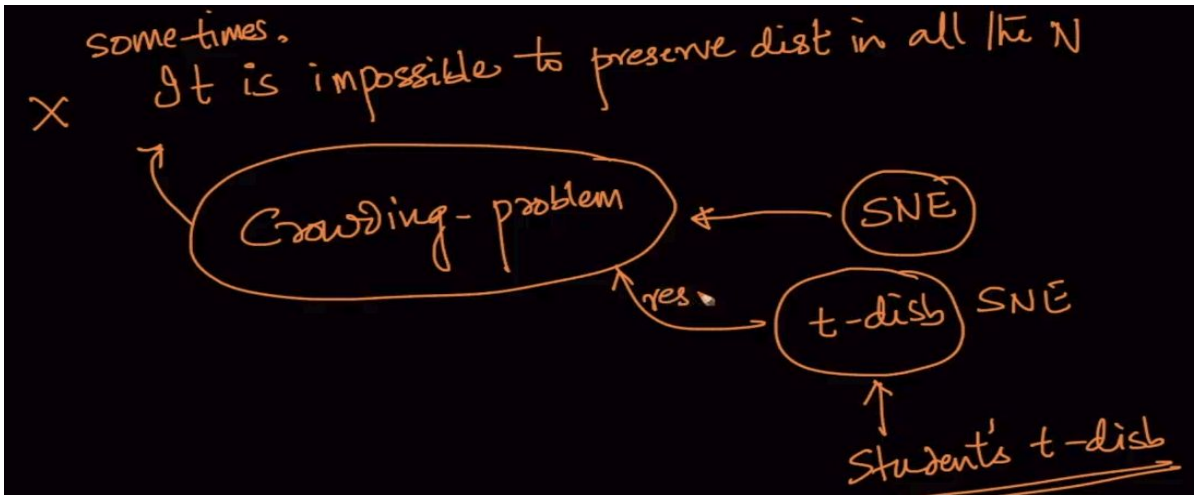
Suppose we have a d-dim data as shown and we are embedding it in 2-d. The neighborhood N(x_1) = $\{x_2, x_3\}$ so when we are looking to embed them we try to preserve their neighborhood as seen above but for the data points which are farther like x_4, x_5 it doesn't guarantee to preserve its distance.

Mathematical Intuition: Fairly advanced math so it is postponed.

CROWDING PROBLEM

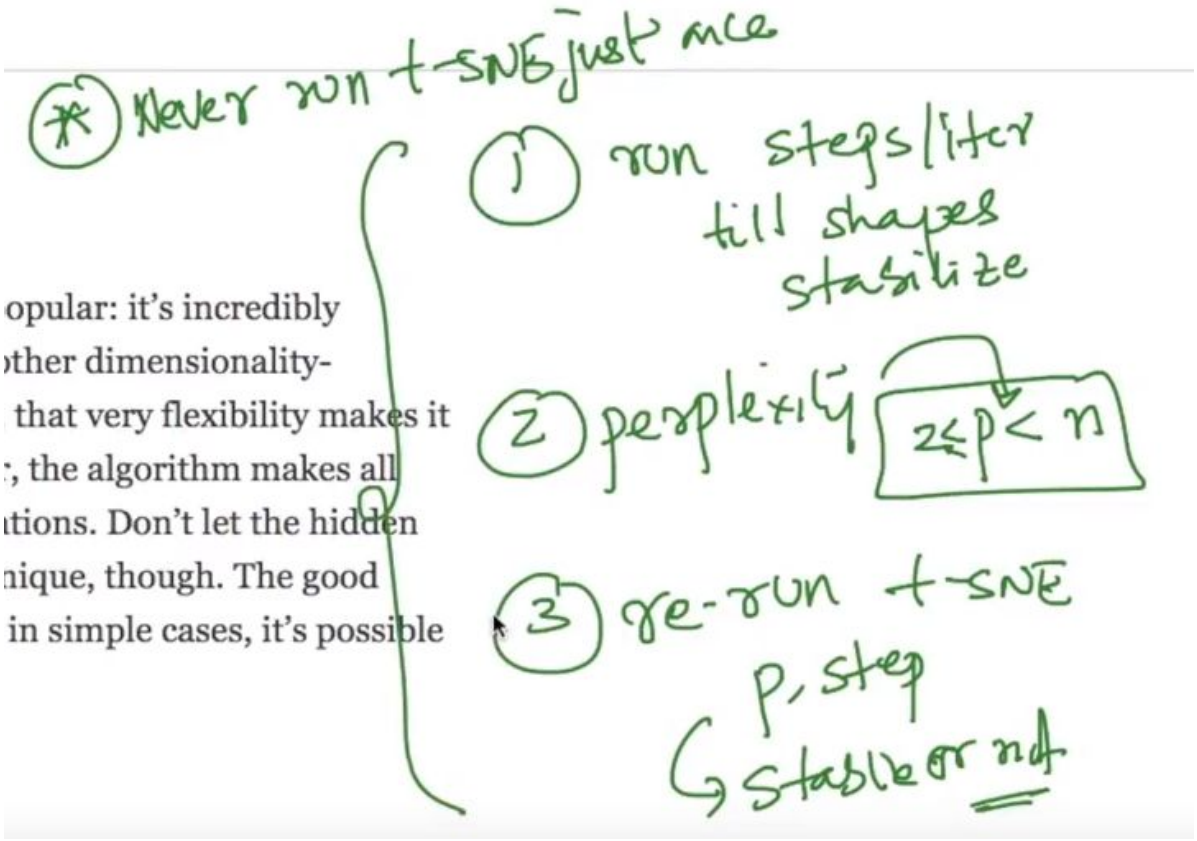


We are trying to embed 2-d to 1-d . So we are placing points to preserve their distances at distance d from x1 as seen N(x1) = { x2,x4} but when doing it the d of x4 becomes greater from x3 and the distance isn't preserved . Same if we try to preserve from x3 . It won't be preserved from x1.



Impossible to preserve distance in all the distances .It is called crowding problem which is solved by t-disb

How to apply t-SNE and interpret its output



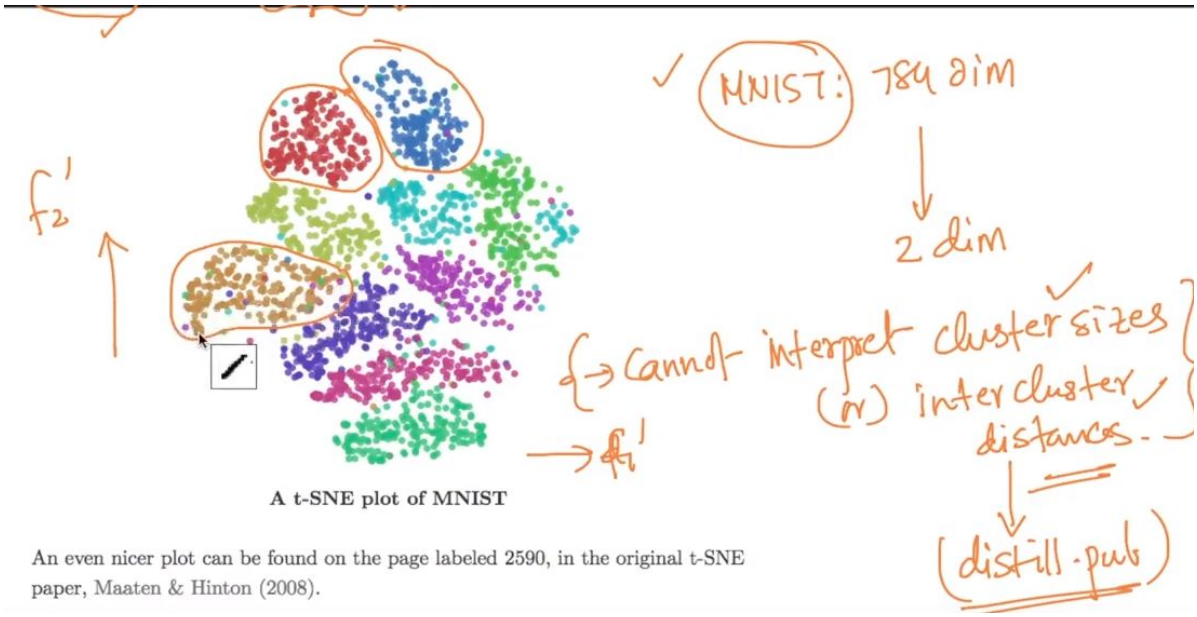
Perplexity : It is nothing but the no. of near data points taken as mentioned

Example: If perplexity is 5 then 5 nearest data points are taken when running t-SNE

[t-SNE visualize video]

[t-SNE visualize blog]

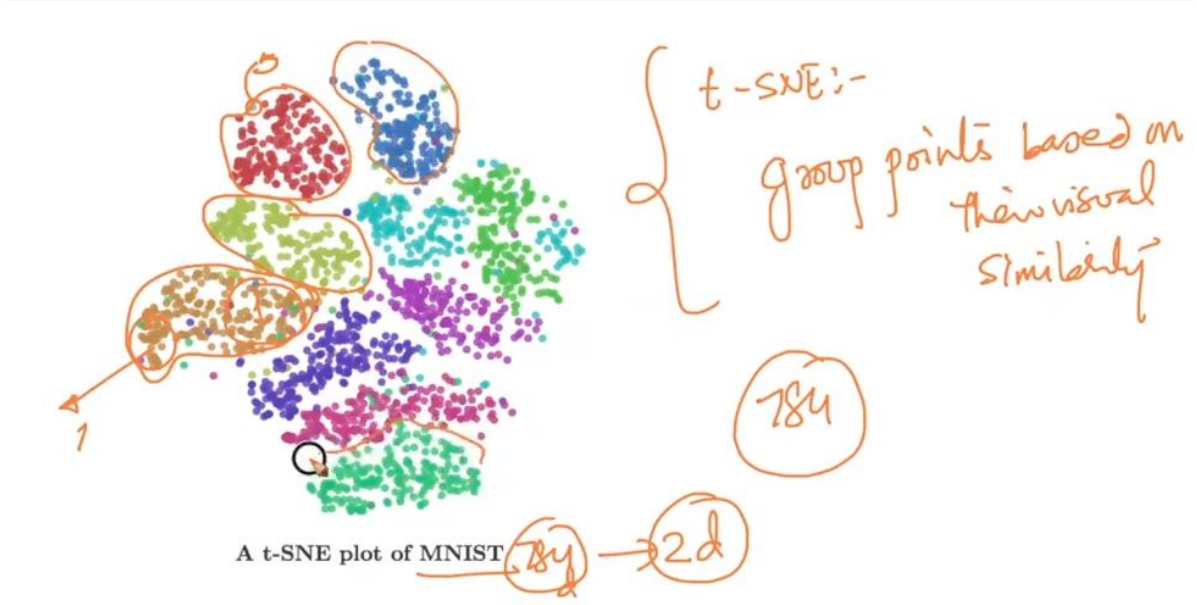
t-SNE on MNIST



A t-SNE plot of MNIST

An even nicer plot can be found on the page labeled 2590, in the original t-SNE paper, Maaten & Hinton (2008).

It is seperating clusters well but we cannot interpret cluster sizes or intercluster distances



A t-SNE plot of MNIST

AFter running t-SNE on MNIST we can see that slanted 1's (*1*) are grouped together and straight 1's are clustered together as well. So it group points based their visual similarity