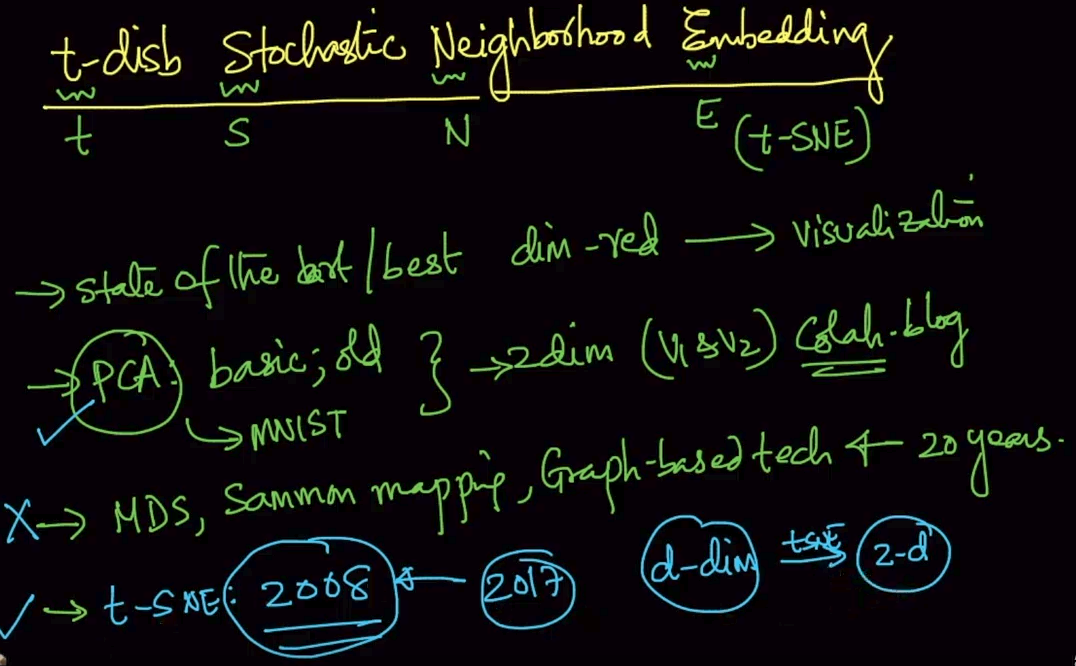
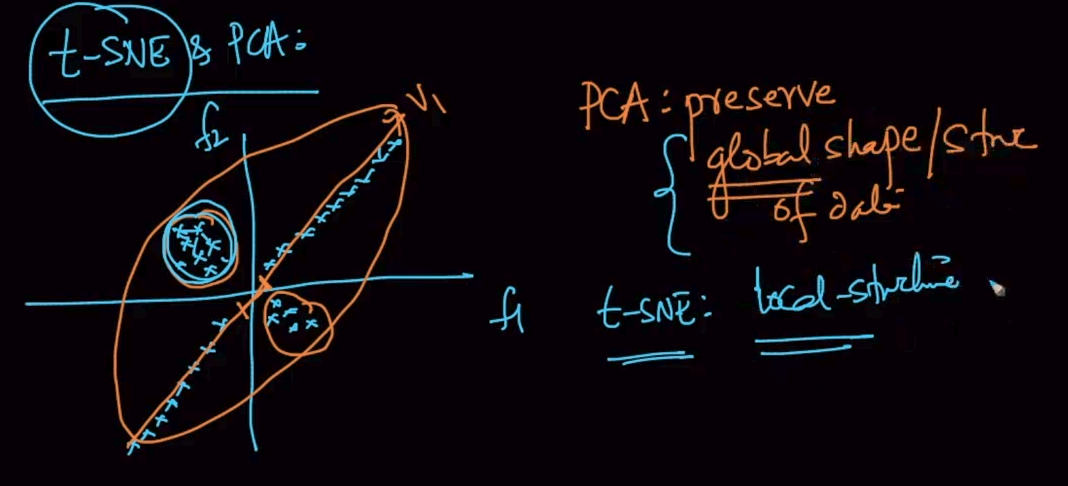
**t-SNE(T-distributed Stochastic Neighborhood Embedding)**

****

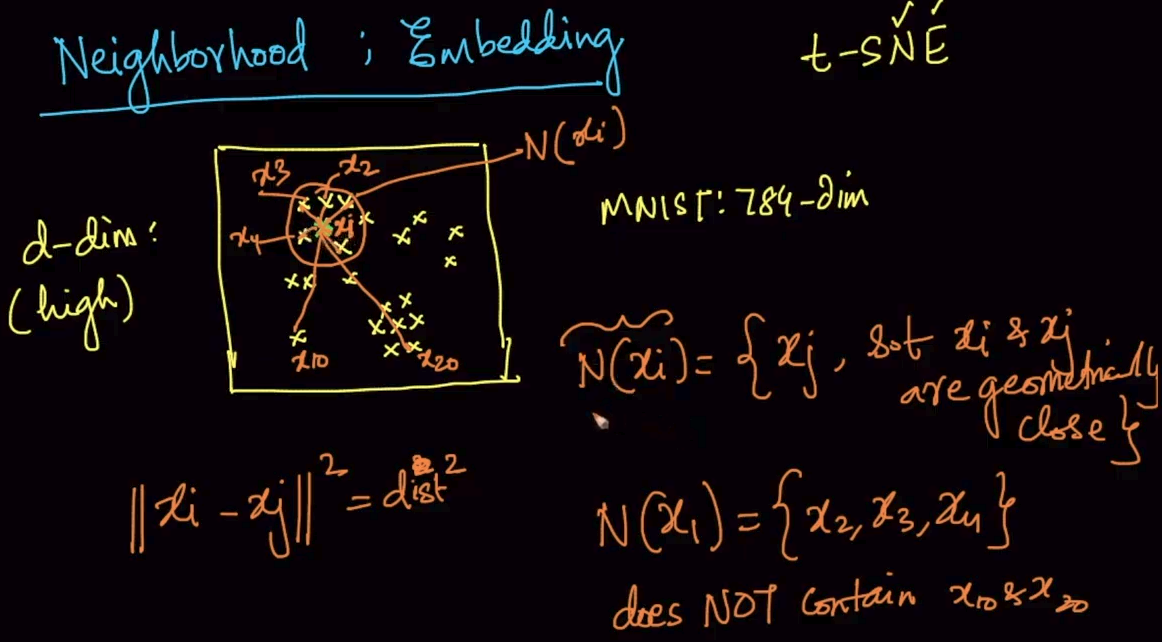
t-SNE is state of the art /best dimensionality reduction technique for visualization.

DIFFERENCE BETWEEN PCA AND t-SNE

**PCA : Preserves global shape/structure of data**

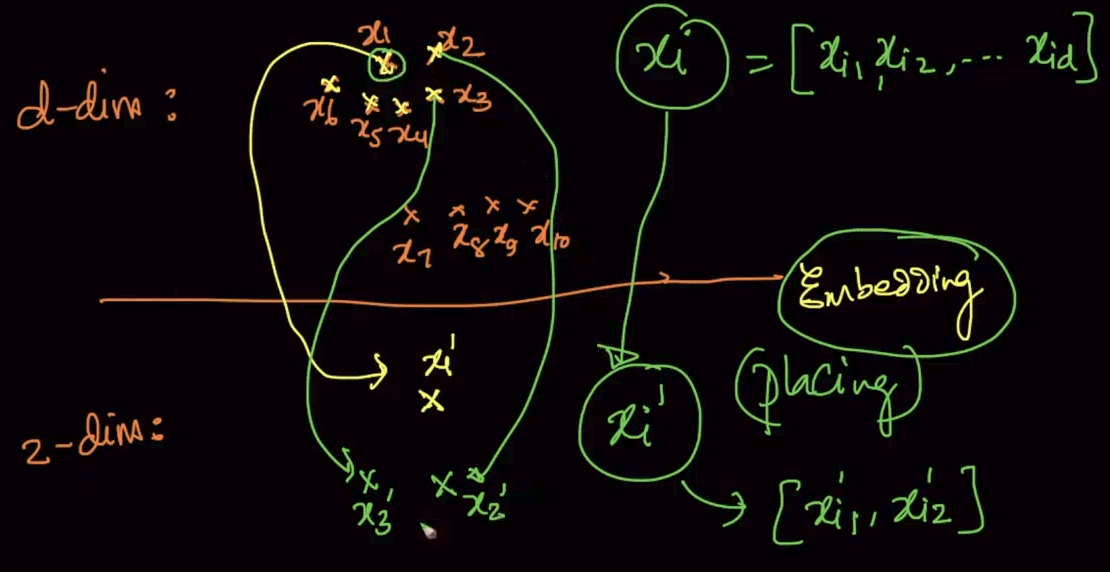
**t-SNE : Local Structure**

**NEIGHBORHOOD**



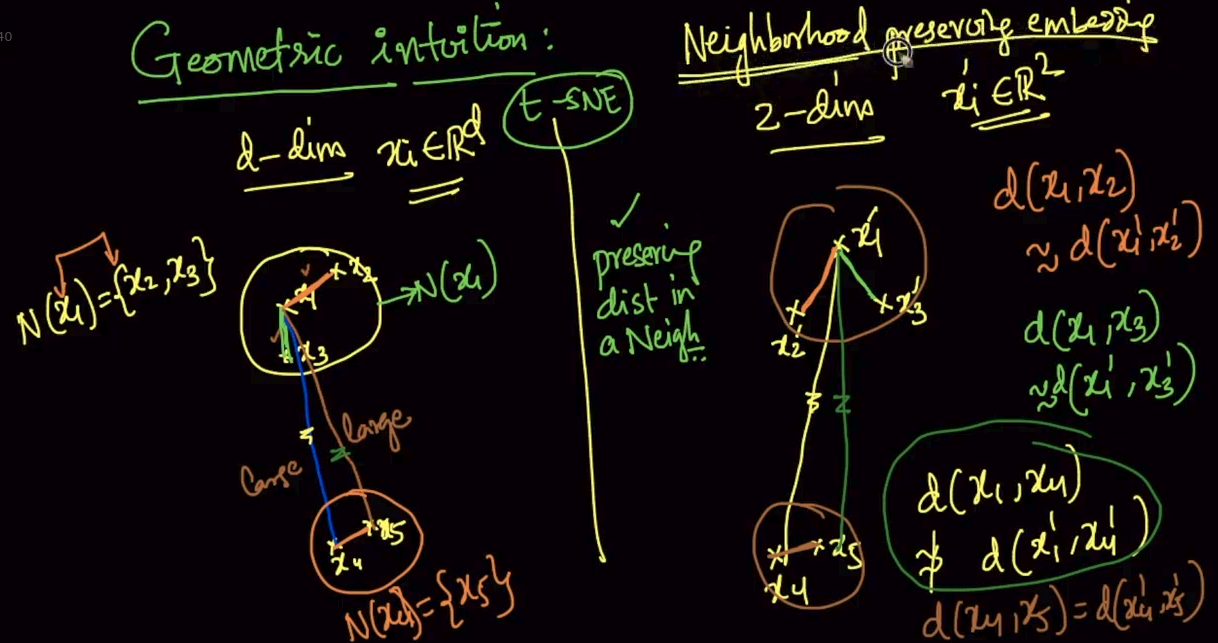
Neighborhood (N) is all the points that are closer to the data point/ Distance between datapoints is small. Here, a data point and it’s N = {,}

**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.

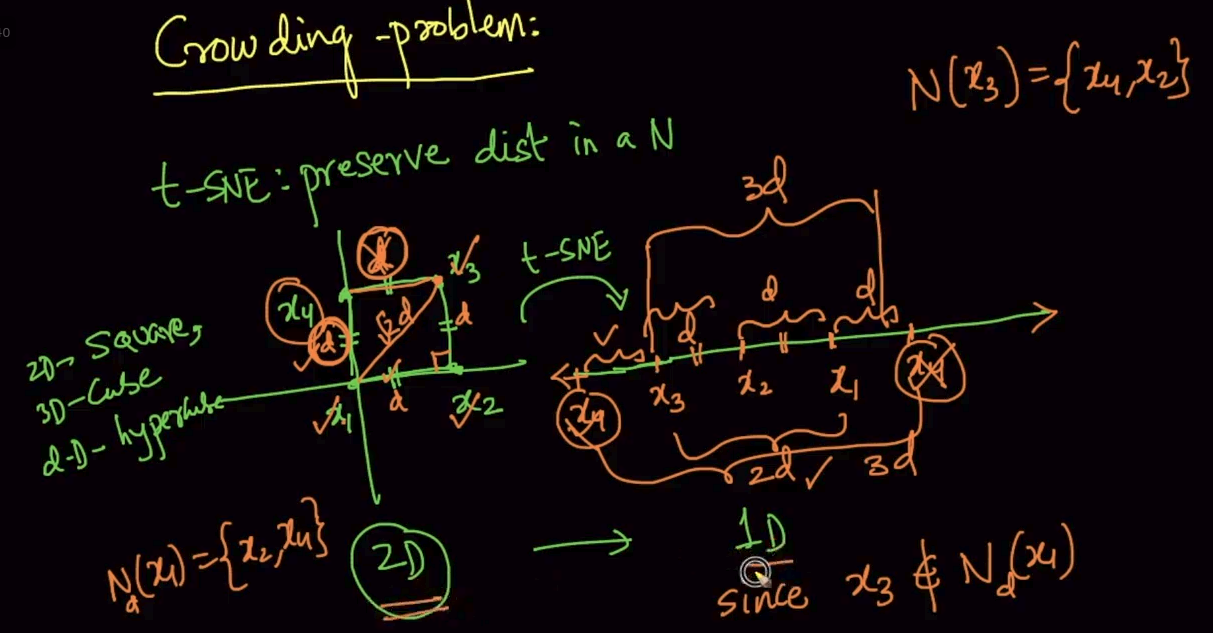
**GEOMETRIC INTUITION of t-SNE**

****

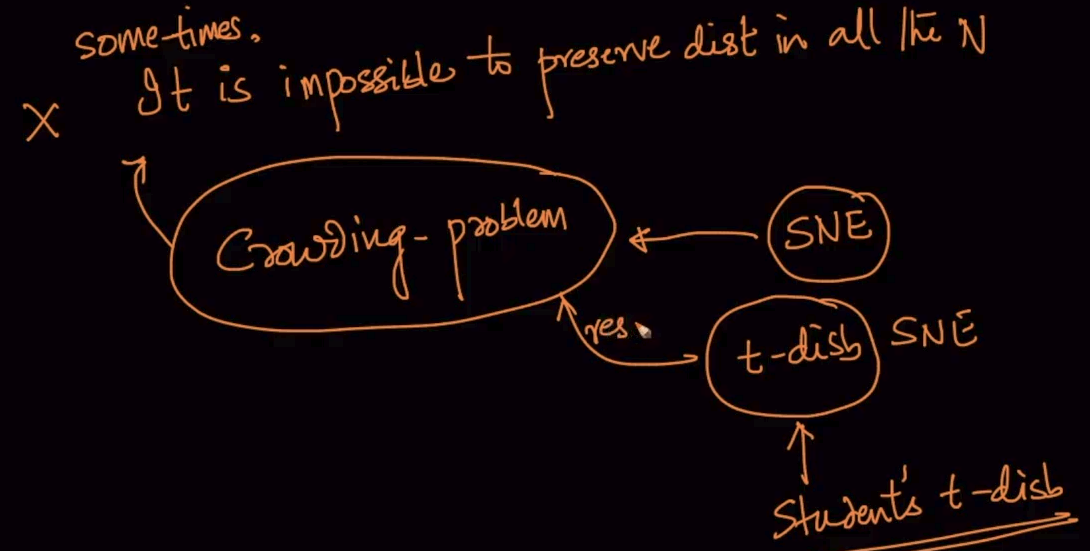
Suppose we have a d-dim data as shown and we are embedding it in 2-d. The neighborhood N() = {} so when we are looking to embed them we try to preserve their neighborhood as seen above but for the data points which are farther like it doesn't guarantee to preserve its distance.

**Mathematical Intuition**: Fairly advanced math so it is postponed.

**CROWDING PROBLEM**

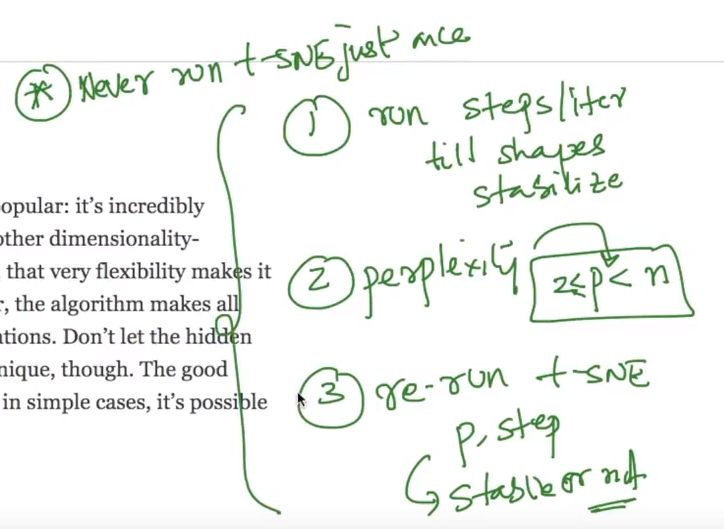
****

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**



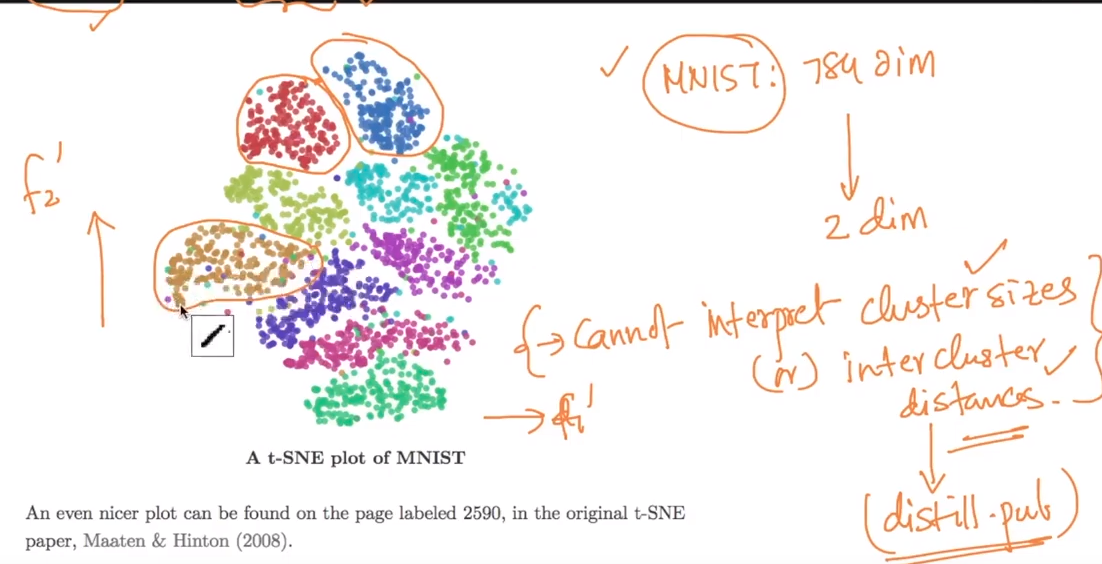
Perplexity : It is nothing but the no. of near data points taken as mentioned

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

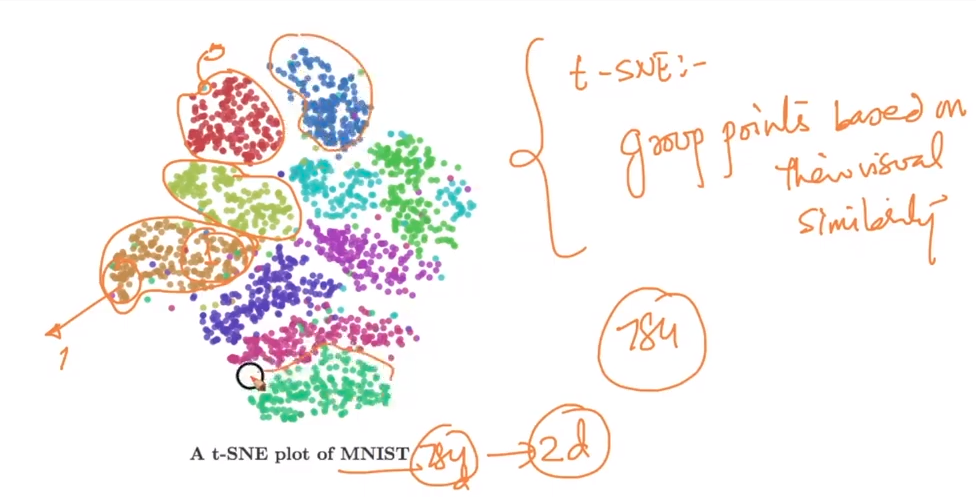
[t-SNE visualize video](https://www.youtube.com/watch?v=PTNfRk0v6YM&list=PLupD_xFct8mHqCkuaXmeXhe0ajNDu0mhZ&index=5)

[t-SNE visualize blog](https://distill.pub/2016/misread-tsne/)

**t-SNE on MNIST**



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



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