t-distb Stochastic Neighborhood Embedding

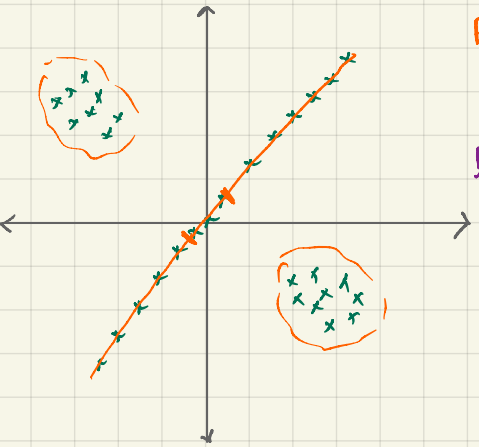
→ state of the art/best dim-red ⟶ visualization

→ PCA : basic ; old → MNIST ⟶ 2 dim (V_1 & V_2)

→ other dim. reduction techniques ⟶ MDS, Sammon Mapping, Graph based tech

→ t-SNE : 2008 d-dim → 2-d

t-SNE & PCA :-

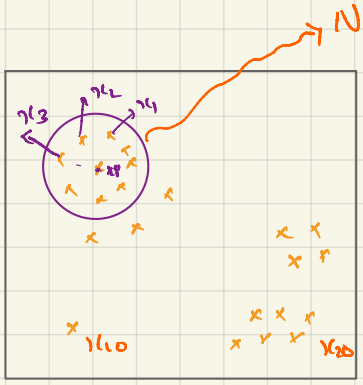


PCA : presserve global shape/structure of data

t-SNE : presserve local shape/structure of data
can choose to

Neighborhood of a point, Embedding

Neighborhood



d-dim (high)

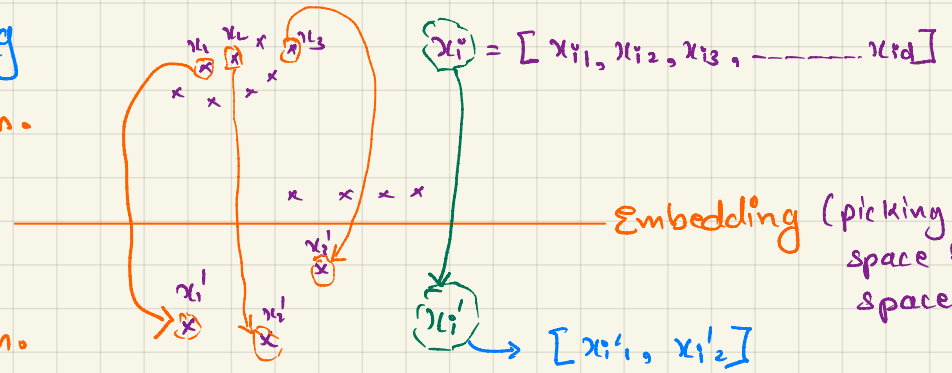
MNIST : 784 -dim

$N(x_i)$ = { x_j , s.t x_i & x_j are geometrically close }

$N(x_1)$ = { x_2, x_3, x_4 }
does not contain x_{10} & x_{20}

$||x_i - x_j||^2 = dist^2$

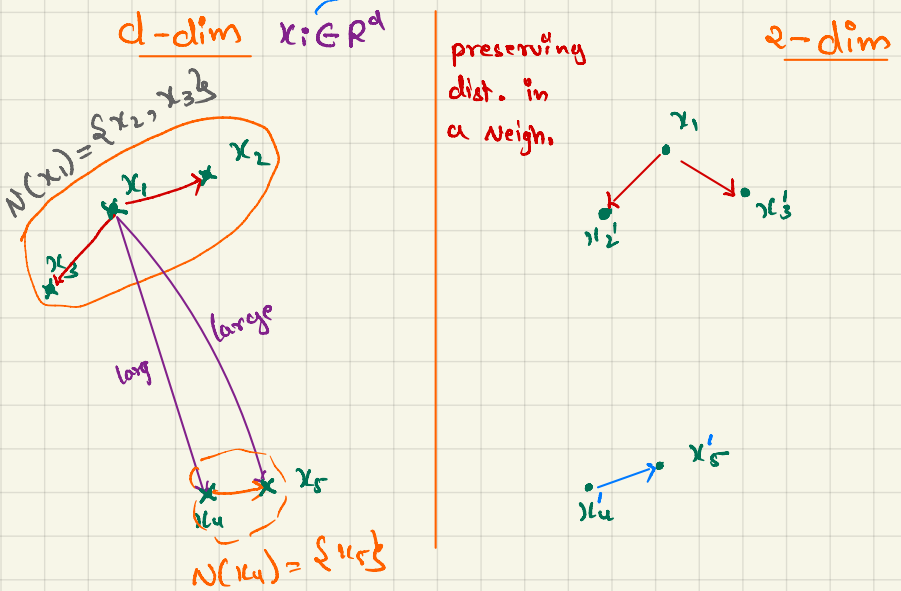
Embedding



Geometric intuition of t-SNE

Geometric intuition:

each x_i' belongs to 2-dimen. here.



$d(x_1, x_2) \approx d(x_1', x_2')$

↳ I want to preserve distance b/w x_1 & x_2 same as x_1' and x_2'.

$d(x_1, x_3) \approx d(x_1', x_3')$

$d(x_1, x_4) \not\approx d(x_1', x_4')$

↳ distance b/w x_1 & x_4 need not to same as x_1' & x_4'.

In 't-SNE' we only preserve neighbourhood distances not the long distances.

$d(x_4, x_5) \approx d(x_4', x_5')$

t-SNE :- Neighborhood preeserving Embedding.

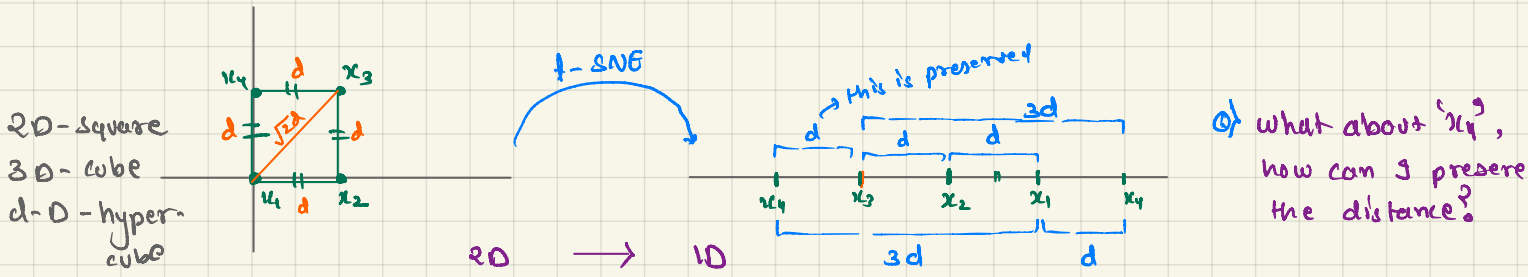
Math-formulation

→ fairly advanced (2008)

→ diverge X

Crowding - problem

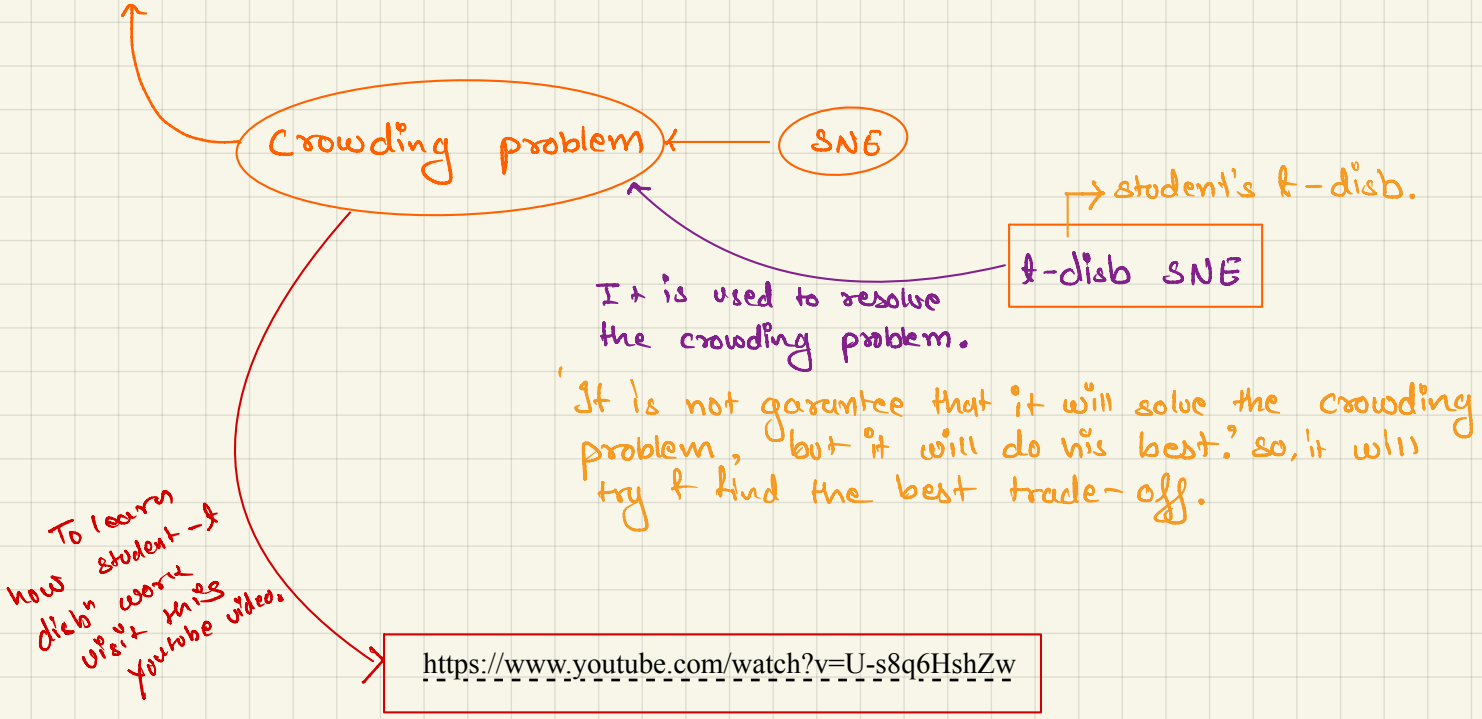
t-SNE : preserve dist in a N



$N_d(x_4) = \{x_2, x_1\}$

$N_d(x_3) = \{x_1, x_2\}$

→ Sometime it is impossible to preserve dist. in all the N.



How to apply t-SNE and interpret its output

→ distill.pub blog (check it)

→ t-SNE is an iterative algorithm. iterative: 1, 2, 3, ... (embedding.)

parameter: step-size = # iteration
1, 2, 3, ... 5000

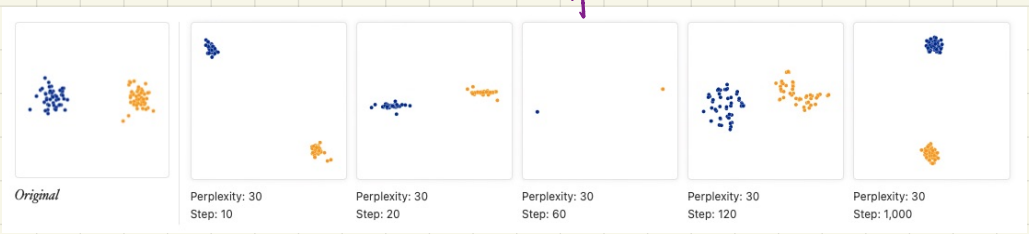
: perplexity → roughly no. of neighbours to whom the distance I want to preserve.

Case 1:- Those hyperparameters really matter

no. of iteration = 5000 (const.)

→ Always run your t-SNE for multiple "perplexity" values to understand the actual shape.

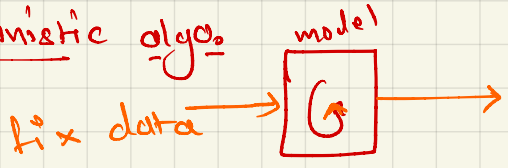
→ if perplexity = # of data-points (n) (always mess)



perplexity = const.

→ Always run your iteration till your shape does not change (stablize).
→ Always run t-SNE multiple times. (coz output may be slightly diff.)
⇒ t-SNE
↳ stochastic
↑
probabilistic [means if you are keeping your 'perplexity' and 'iteration' same but, still you could get slightly different result. That is because t-SNE is not a deterministic algorithm.]

deterministic algo.

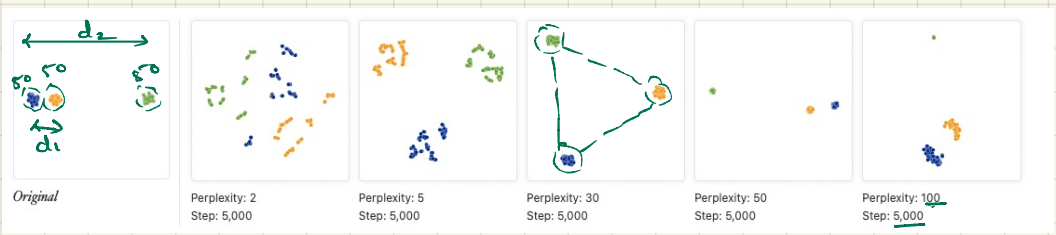


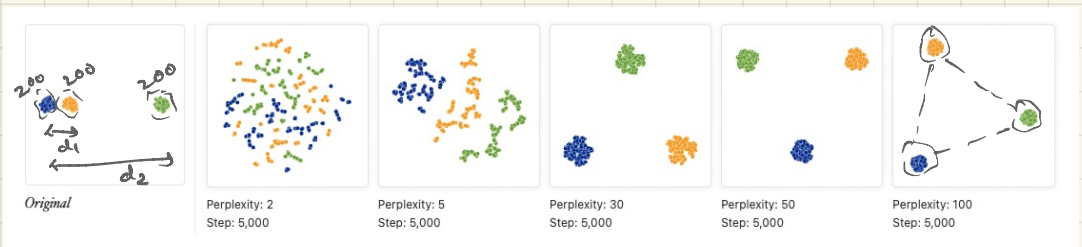
Case II :- Cluster size in a t-SNE plot mean nothing

t-SNE :-
- expand dense cluster.
- shrinks sparse cluster.

→ You cannot get cluster sizes or you can't say your cluster have same density.
(can't say anything about cluster size)

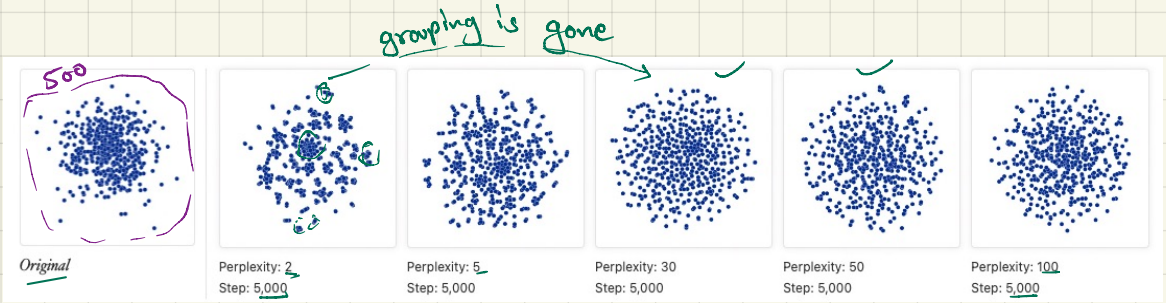
Case - III : Distances b/w clusters might not mean anything.





→ t-SNE does not preserve distance b/w cluster.

Case: IV Random noise doesn't always look random.



→ Never conclude that there is structure in data by just looking one - perplexity. Always look at more perplexity value.