t-SNE :
t-distributed Stochastic Neighbor Embedding

12.15.2021

Hannah Do

○. Introduction

Popular method for **exploring high-dimensional data**, introduced by Laurens van der Maaten and Geoffrey Hinton in 2008.

PCA uses covariance and linear transformation on matrices to reduce dimensions mainly on ***Linear Data***.

However, t-SNE can create lower-dimensonal “maps” from data with hundreds or even thousands of dimensions, including ***Non-linear Data.***

1. Short Summary

t-distributed Stochastic Neighbor Embedding (SNE)

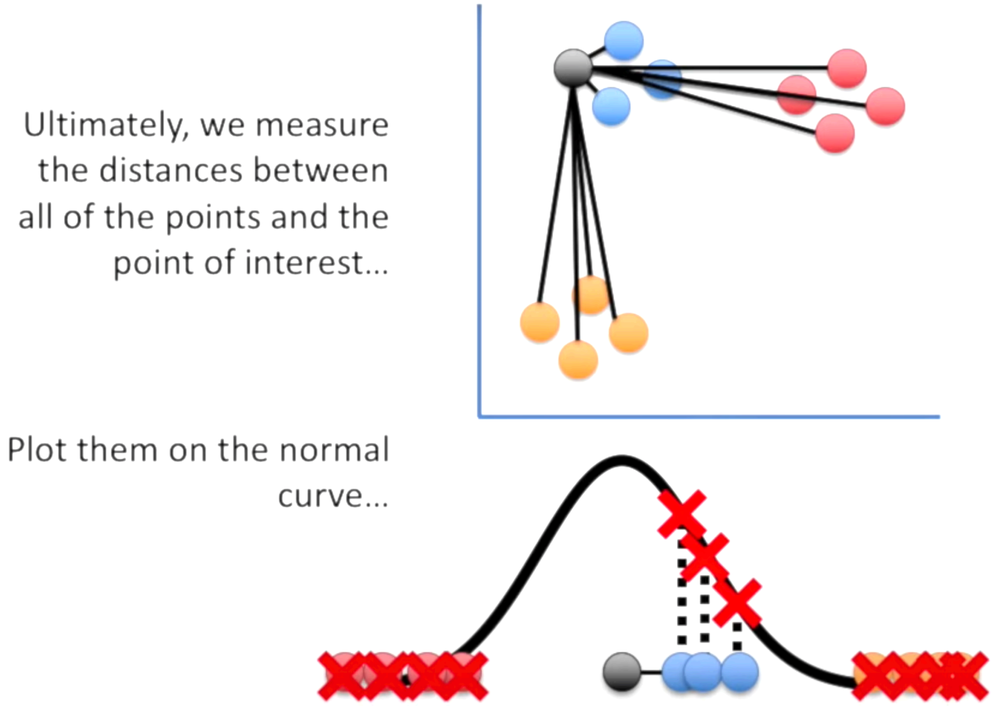
A. Calculates Euclidean distances between data points

B. Computes Conditional probabilities known as *similarity scores*.

C. Similarity score from lower (projected) dimension is compared with the similarity score from the higher (original) dimension until the difference is reduced to zero.

2. How does it work?

1. Find the distance between the target point and the rest of the data points & plot them on the normal curve

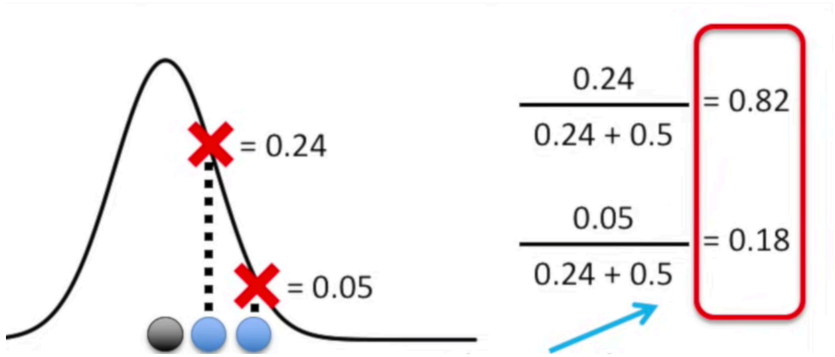


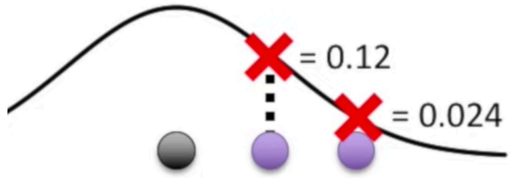
Statquest: t-SNE, Clearly Explained
https://www.youtube.com/watch?v=NEaUSP4YerM

2. How does it work?

2. Compute the **similarity score** by dividing the score from sum of all scores

$$\frac{\text{Score}}{\text{Sum of all scores}} = \text{Scaled Score}$$





$$Perp(P_i) = 2^{H(P_i)},$$

$$H(P_i) = -\sum_j p_{j|i} \log_2 p_{j|i}.$$

t-SNE's special parameter - **perplexity** - determines *the radius of the Gaussian distribution.*

Big radius leads to high entropy for the distribution over neighbors of i, whereas small radius leads to low entropy.

Performance of SNE is fairly robust to changes in the perplexity **ranging from 5 to 50**

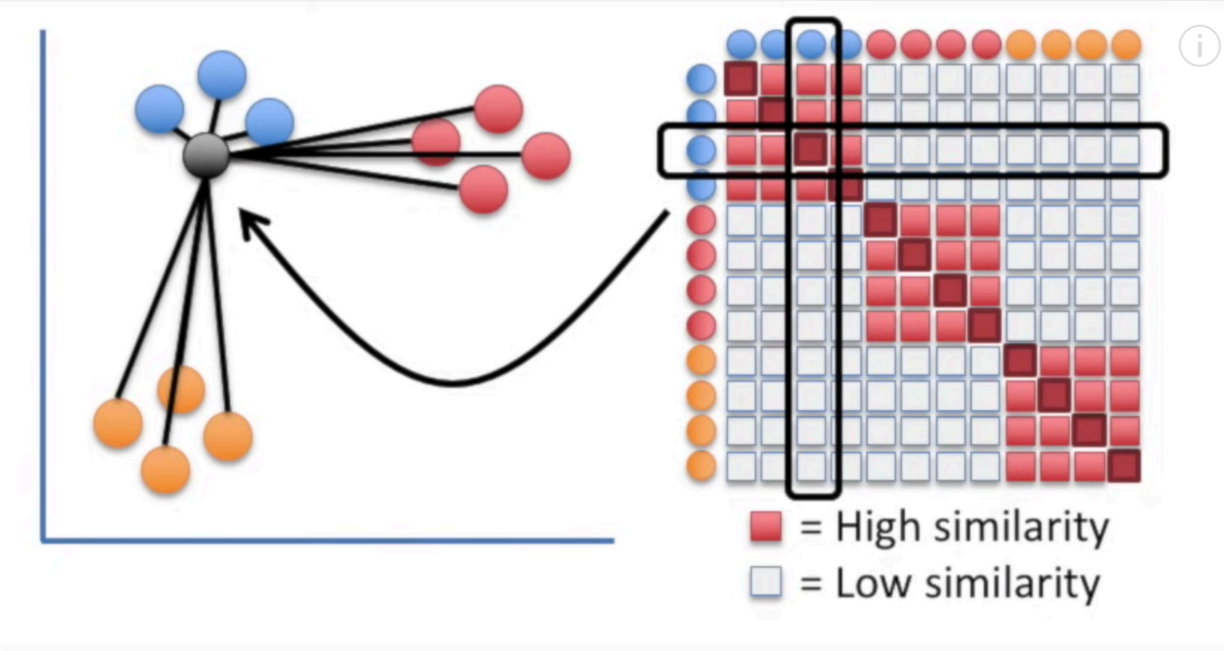
Statquest: t-SNE, Clearly Explained
https://www.youtube.com/watch?v=NEaUSP4YerM

2. How does it work?

3. Compute the **similarity score matrix** by getting *average* of the similarity score from both directions

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n},$$

Similarity of **data point i and j** as a conditional probability



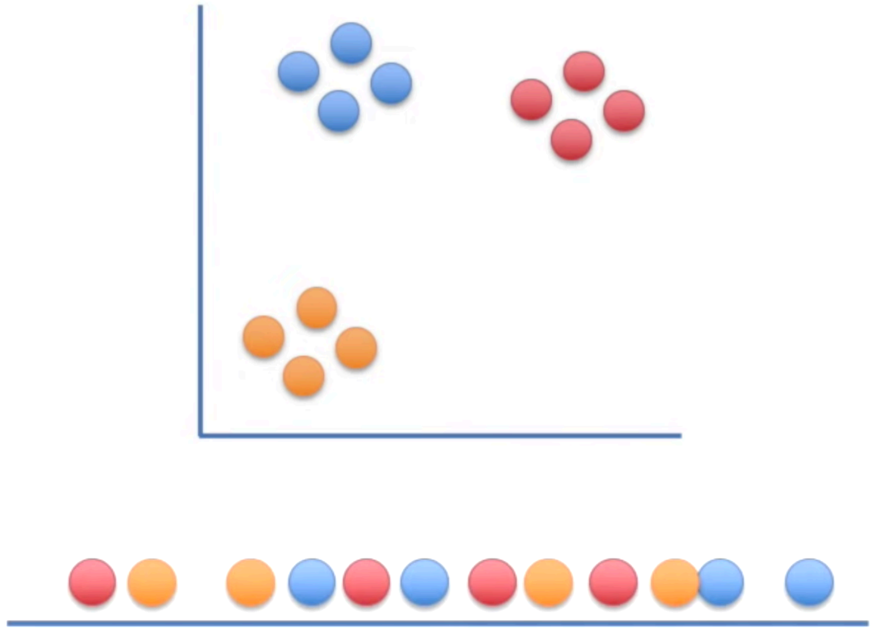
Statquest: t-SNE, Clearly Explained
https://www.youtube.com/watch?v=NEaUSP4YerM

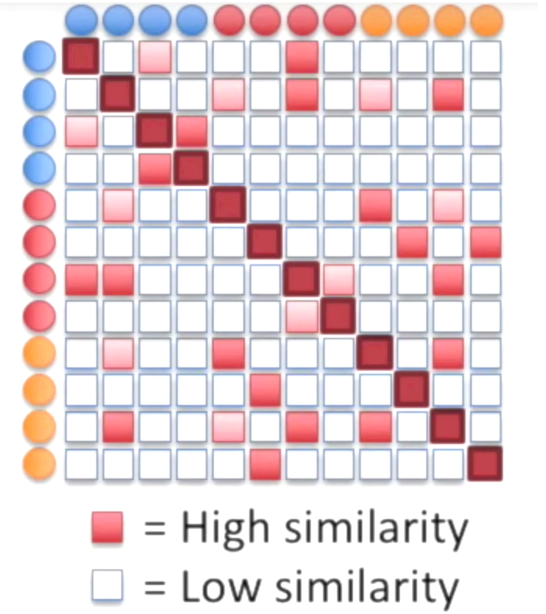
2. How does it work?

4. Randomly project the data onto the lower dimension and create a similarity score matrix

$$q_{ij} = \frac{(1 + |y_i - y_j|^2)^{-1}}{\sum_{k \neq l} (1 + |y_k - y_l|^2)^{-1}}$$

Using **t-distribution** for the similarity score of the projected data points to avoid crowding problem

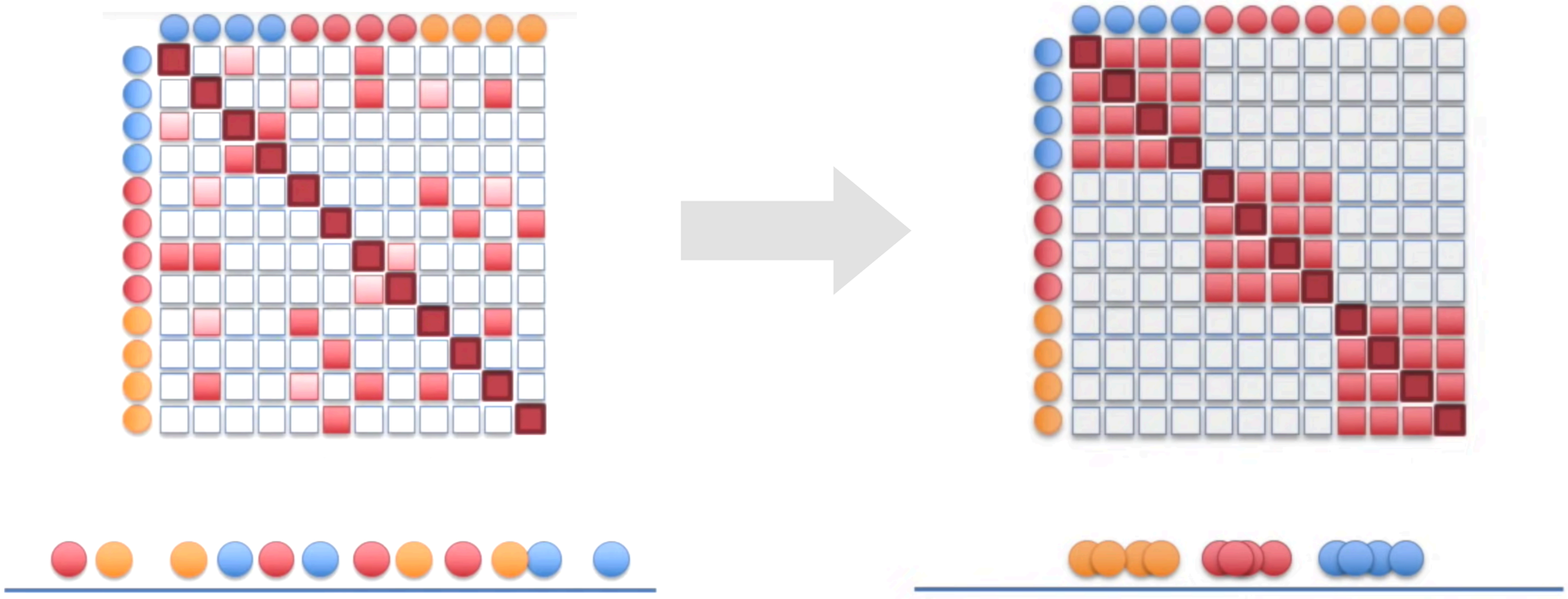




Statquest: t-SNE, Clearly Explained
https://www.youtube.com/watch?v=NEaUSP4YerM

2. How does it work?

5. Compare the new similarity score with the previous similarity score from the higher dimension and reduce the difference by moving each point one at a time



Statquest: t-SNE, Clearly Explained
https://www.youtube.com/watch?v=NEaUSP4YerM

3. Math behind t-SNE

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n},$$

P_i : Conditional Probability distribution over all other data points given data point x_i

$$q_{ij} = \frac{(1 + |y_i - y_j|^2)^{-1}}{\sum_{k \neq l} (1 + |y_k - y_l|^2)^{-1}}$$

Q_i : Conditional probability distribution over all other map points given map point y_i.

Means of a student-t distribution with one degree of freedom.

$$C = KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}.$$

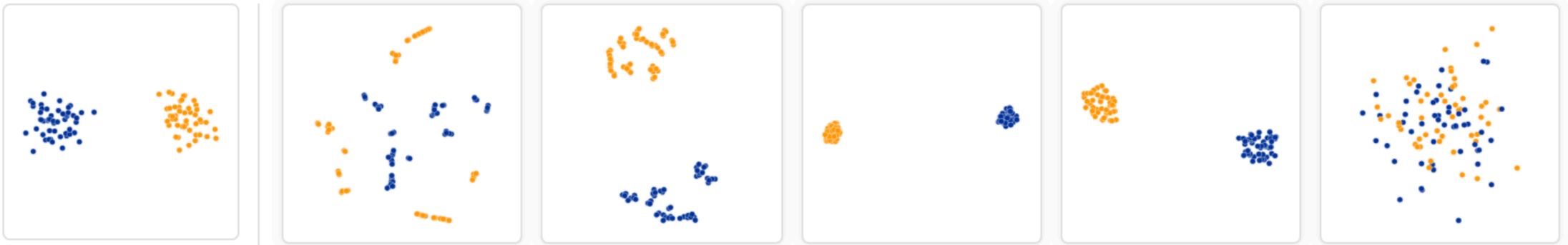
Cost function (Single Kullback-Leibler divergence) to minimize the difference between P and Q
(Symmetrized version of the SNE cost function)

$$\frac{\delta C}{\delta y_i} = 4 \sum_j (p_{ij} - q_{ij})(y_i - y_j) \left(1 + \|y_i - y_j\|^2\right)^{-1}$$

Gradient function to reduce the cost function

4. t-SNE Exploration Tool

Distill : How to use t-SNE Effectively



Original

Perplexity: 2
Step: 5,000

Perplexity: 5
Step: 5,000

Perplexity: 30
Step: 5,000

Perplexity: 50
Step: 5,000

Perplexity: 100
Step: 5,000

https://distill.pub/2016/misread-tsne/

1. References

1. Visualizing Data using t-SNE, Laurens van der Maaten, Geoffrey Hinton
   Journal of Machine Learning Research 9 (2008) 2579-2605

2. Statquest: t-SNE, Clearly Explained
   https://www.youtube.com/watch?v=NEaUSP4YerM

3. Distill : How to Use t-SNE Effectively
   https://distill.pub/2016/misread-tsne/