t-SNE Clearly Explained

- t-SNE does is it takes a high-dimensional data set and reduces it to a low-dimensional graph ***that retains a lot of the original information***.
- What **t-SNE does is find a way to project data** into a low dimensional space so that the **clustering** in the high dimensional space **is preserved**.

☐ After putting the points on the number line in a random order, t-SNE moves these points, a little bit at a time, until it has clustered them.

➢ Determine the "similarity" of all the points in the scatter plot.

○ Random Value를 Normal Curve의 중심으로 하여 모든 값 간의 Distance를 측정하여 line 즉, unscaled 된 상태의 Normal Distribution을 기록합니다.

여기서 *Normal Distribution*은

i. 상대적으로 먼 거리에 위치한 값일수록 *low similarity* 임을 나타냅니다

ii. 각 정규분포의 너비*(Width)*는 거리의 밀도*(Density)*를 나타냅니다.

iii. ii 와 같은 이유로 각 *Disturibution*을 나타내는 *curve*의 *width* 가 달라지게 되고,

iv. 각 ***Cluster***에 대하여 동일한 ***Distribbution***내에서의 상대적인 비교를 위해**,** *Similarity Scores*에 대한 *Scailing*이 필요합니다.

v. *MinMax Scaling*과 같은 방법으로 *Scailing*할 수 있습니다.

a. t-SNE gas a "perplexity" parameter equal to the expected density, and that comes into play, **but the clusters are still more "similar" than you might expect.**

vi. *by using t-distribution, we* ***calculate "unscaled" similarity scores for all the points*** *and then scale them like before.*

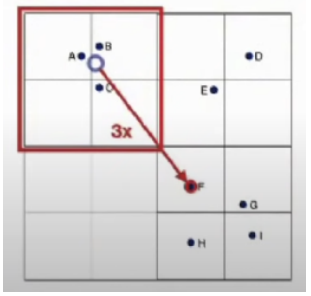
vii. ***Gradient interpretation***

$$\frac{\partial C}{\partial \mathbf{y}_i} = 4\sum_{j \neq i}(p_{ij} - q_{ij})(1 + \|\mathbf{y}_i - \mathbf{y}_j\|^2)^{-1}(\mathbf{y}_i - \mathbf{y}_j)$$

a. ***yi - yj***는 ***spring***을 나타내고**,** ***pij - qij***는 각각 ***high dimensional space***와 ***low dimensional space***를 나타내어 ***extraction***과 나머지를 통해 ***compression***을 진행시킨다**.**

b. ***what this sum is basically doing is taking all forces that act on the single point.***

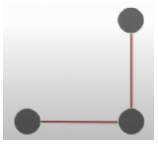
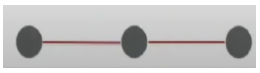
★ ***Barnes-Hut-SNE with Barness-Hut-approximation***



☑ ~~***Why the "t-distribution" is used***~~

: Suppose data is intrinsically high-dimensional

: We try to model the local structure of this data in the map

:  

- **dissimilar points have to be modeled as too far apart in the map**
  - **red line is the small distances so the local structure**
  - **the distance between the two other, it's large pairwise structure so the global structure**
- **embed data in 1D while preserving local structure**
- **the distance between the two points that are far away have grown.**
- **Now, by using this heavy tailed distribution, it is under the Gaussian, it gives me a density of 1**
- **So for dissimilar points, these heavy tailed qij is basically allow dissimilar points to be modeled too far apart in the map.**