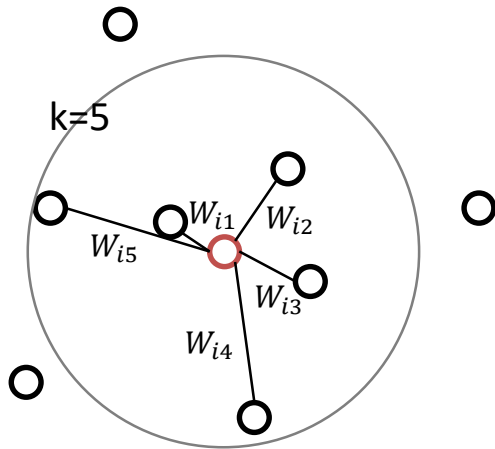


Stochastic Neighbor Embedding

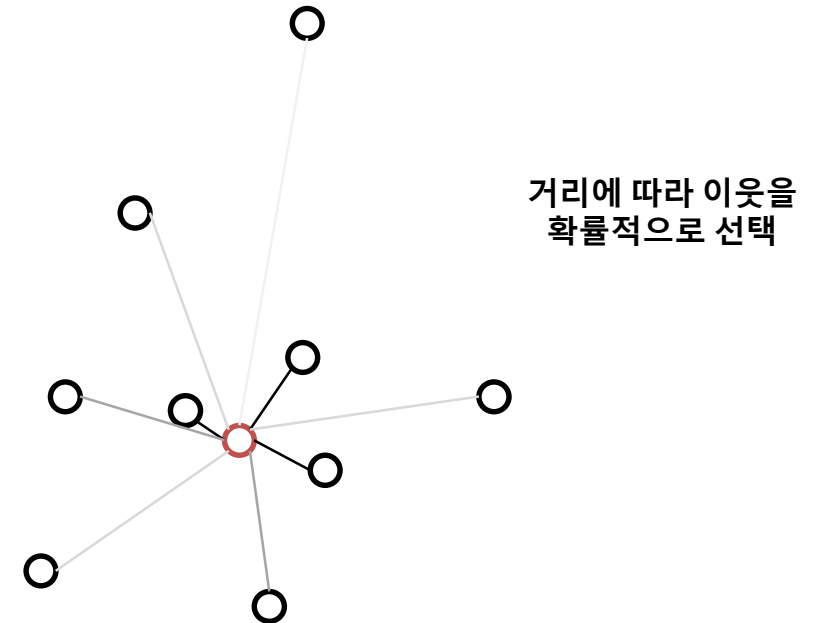
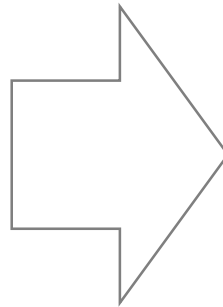
Methods

❖ LLE vs SNE

- LLE는 이웃의 개수 k 를 정하고 이웃이 변하지 않음
- 하지만 실제로 이웃보다 더 먼 다른 관측치도 x_i 를 표현할 수 있음
- 이웃의 선택을 거리에 따라 확률적으로 선택하도록 변경



Deterministic Neighbor (LLE)



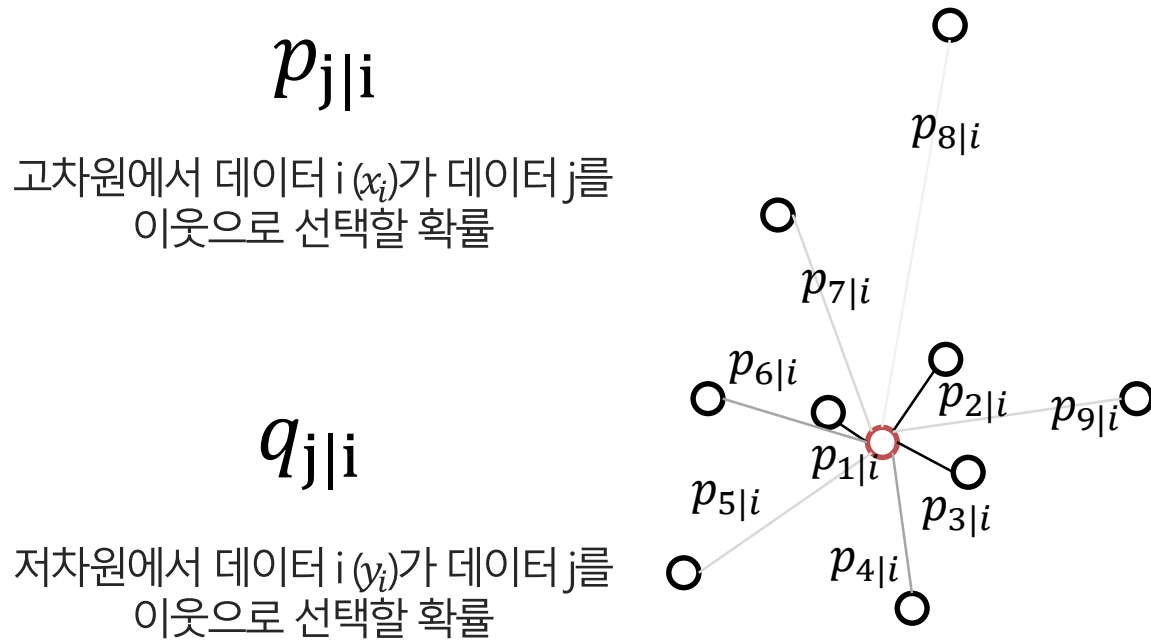
Stochastic Neighbor (SNE)

Stochastic Neighbor Embedding

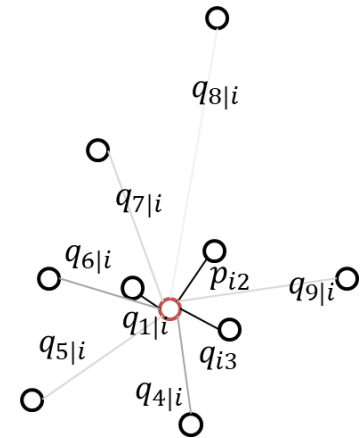
Methods

❖ LLE vs SNE

- LLE는 이웃으로 표현되는 선형결합을 유지하도록 임베딩
- SNE는 이웃으로 선택할 확률 분포를 유지하도록 임베딩



확률분포 보존
= 데이터 핵심 구조 보존



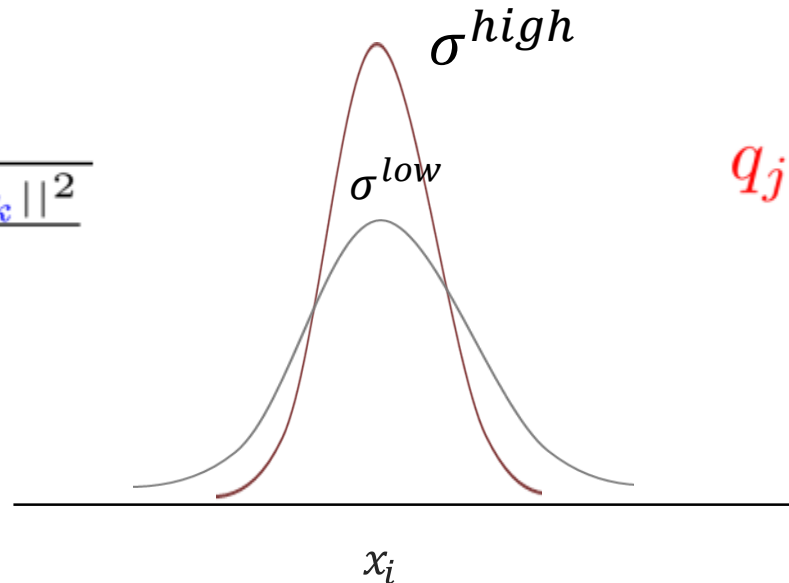
Stochastic Neighbor Embedding

Methods

❖ Neighbor selection probability

- 이 때, x_j 가 x_i 의 이웃으로 선택될 확률 $p_{j|i}$ 는 x_i 와 x_j 의 거리에 따라 줄어들어야 함
- σ 를 통해서 고차원에서 가장 x_i 를 잘 표현할 수 있는 확률분포 $p_{j|i}$ 를 탐색

$$p_{j|i} = \frac{e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma_i^2}}}{\sum_{k \neq i} e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_k\|^2}{2\sigma_i^2}}}$$



$$q_{j|i} = \frac{e^{-\|\mathbf{y}_i - \mathbf{y}_j\|^2}}{\sum_{k \neq i} e^{-\|\mathbf{y}_i - \mathbf{y}_k\|^2}}$$

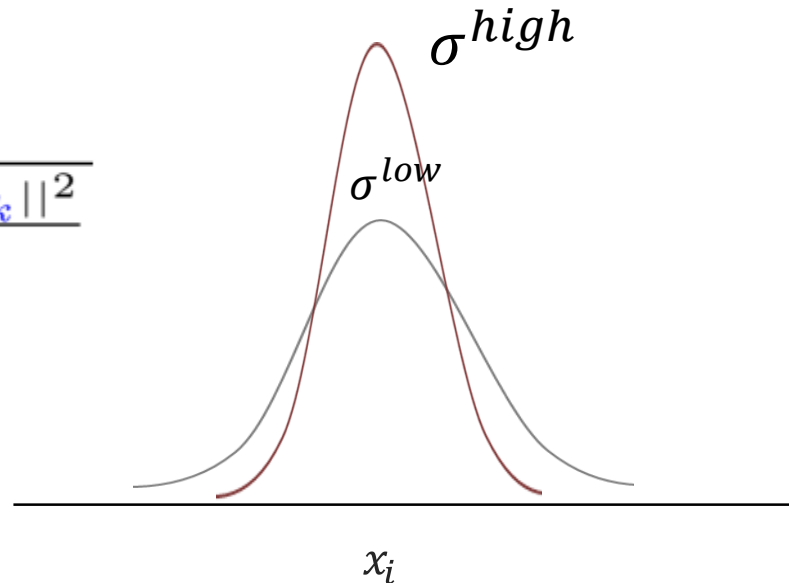
Stochastic Neighbor Embedding

Methods

❖ Neighbor selection probability

- σ_i 를 통해서 고차원에서 가장 x_i 를 잘 표현할 수 있는 확률분포 $p_{j|i}$ 를 탐색
- Perplexity를 정의하여 지정한 perplexity를 만족하는 σ_i 탐색
- Perplexity는 보통 5~50으로 설정하며 SNE의 성능은 perplexity에 강건함

$$p_{j|i} = \frac{e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma_i^2}}}{\sum_{k \neq i} e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_k\|^2}{2\sigma_i^2}}}$$



$$\text{Perplexity}(P_i) = 2^{H(P_i)}$$

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

Stochastic Neighbor Embedding

Methods

❖ Objective Function

- 고차원의 확률분포 $P(= \sum_j p_{j|i})$ 와 저차원의 확률분포 $Q(= \sum_j q_{j|i})$ 의 차이가 줄어들도록 학습
→ 모든 관측치 x_i 에 대하여 P 와 Q 의 KL-divergence가 최소화되도록 학습

$$\mathcal{C} = \sum_i KL(P_i || Q_i) = \sum_i \sum_j p_{j|i} \log\left(\frac{p_{j|i}}{q_{j|i}}\right)$$

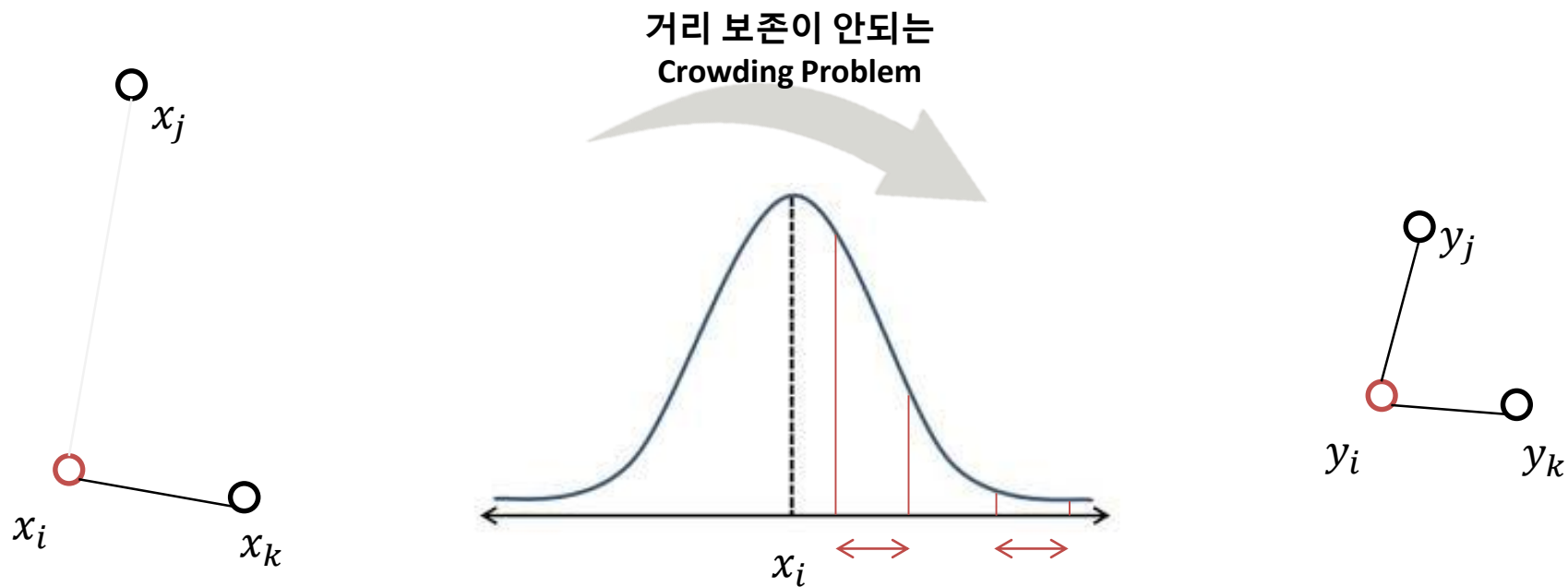
$$\frac{\partial \mathcal{C}}{\partial y_i} = 2 \sum_j (y_j - y_i)(p_{j|i} - q_{j|i} + p_{i|j} - q_{i|j})$$

t-SNE

Methods

❖ Crowding Problem

- 정규 분포를 사용해 $q_{j|i}$ 를 구하면 적당히 떨어진 이웃 j 와 많이 떨어진 이웃 k 가 선택될 확률의 차이가 크지 않음
→ 저차원에서 이웃과의 거리 차이를 보존할 수 없고 뭉치게 되는 crowding problem 발생

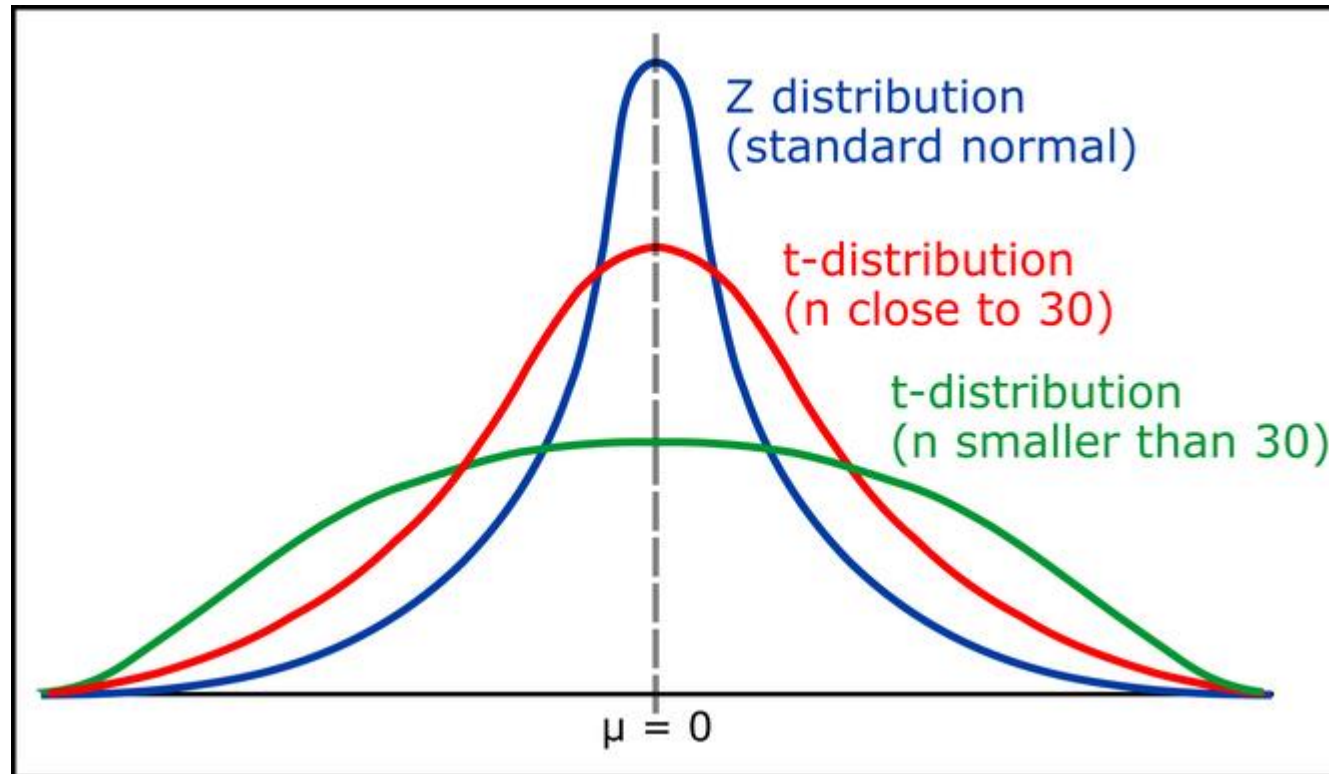


t-SNE

Methods

❖ T-SNE: t-distribution instead of normal distribution

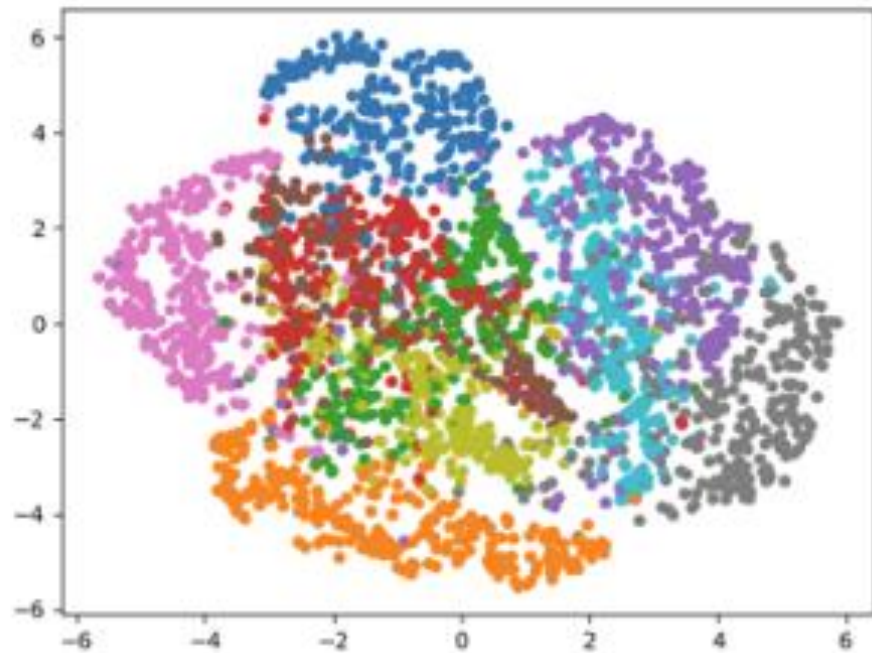
- 꼬리가 좀 더 넓은 t-distribution을 활용하여 crowding problem 해결



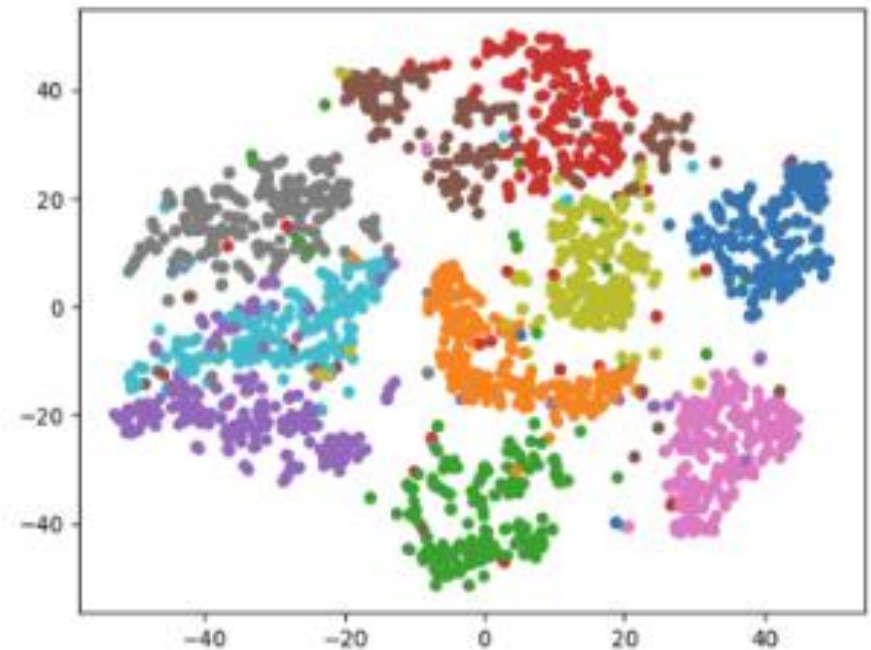
t-SNE

Methods

❖ t-SNE vs SNE (MNIST dataset)



SNE



t-SNE