

t-SNE

강사 : 백병인

pi.paek@modulabs.co.kr

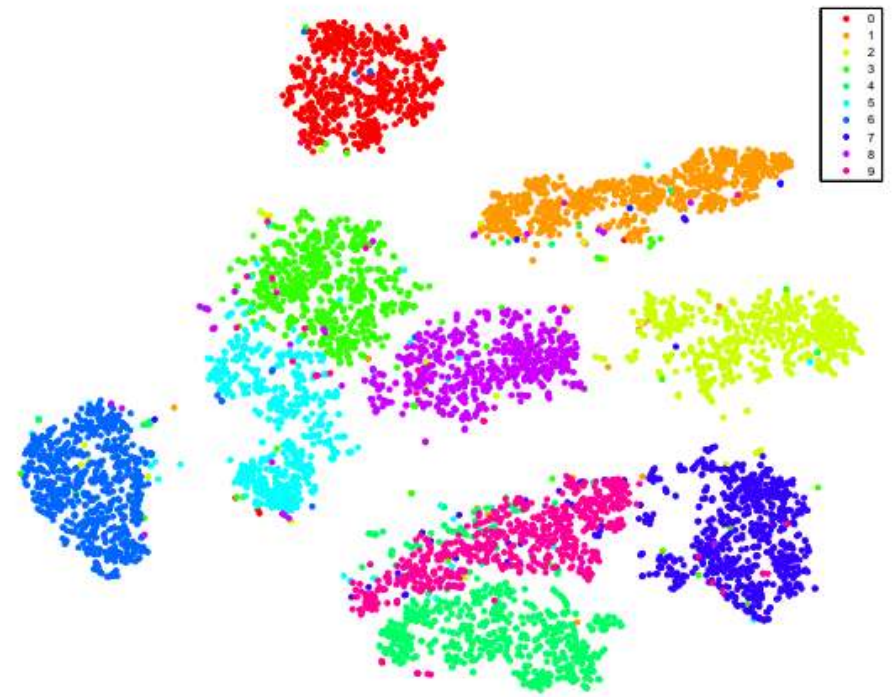
모두의연구소 Research Scientist



2019 모두의연구소

Visualization of Features

- 좋은 Feature는 class간 명확하게 구별되는 속성이어야 한다.
- 그러나 Feature Vector는 일반적으로 수십 dimension이나 된다. 좋은 Feature가 구해졌는지 확인할 방법이 있을까?
- Feature Visualization
 - 다차원의 Feature Vector를 저차원(2~3차원)으로 차원 축소하여 좌표면 상에 시각화
 - Feature Vector의 분포가 class별로 잘 clusterin되어 있는지 시각적으로 확인하는 방법
 - PCA, t-SNE



(a) Visualization by t-SNE.

Feature Dimension Reduction

- 하고싶은 것
 - N개의 고차원 feature vector x_i 를
 - N개의 저차원 feature vector y_i 로 투영하기 위해
 - x_i 와 x_j 의 거리가 가깝다면
 - y_i 와 y_j 의 거리도 가깝도록 y_i 를 y_j 방향으로 이동



t-SNE (1)

- t-SNE(Stochastic Neighbor Embedding)
 - 확률적인(Stochastic) 이웃(neighbor) 뽑기를 통한 Embedding 방법
 - x_i 가 어떤 x_j 를 자기와의 거리에 대한 Gaussian 분포 확률에 따라 자기 이웃으로 뽑는다고 하자.
- y_i 와 y_j 에 대해서도 이웃 뽑기 확률이 똑같도록 하면 되지 않을까?
- 그러려면 p_{ij} 와 q_{ij} 의 분포가 유사해지도록 하면 되겠다.

$$p_{ij} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma^2)}{\sum_{k \neq l} \exp(-\|x_k - x_l\|^2 / 2\sigma^2)}$$

단, q의 분포는 Gaussian이 아니라 t-분포로 표현했다. 그래서 t-SNE !!

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

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



t-SNE (2)

- t-SNE(Stochastic Neighbor Embedding)
 - 학습해야 할 것은 y_i 의 임베딩 벡터.
 - 그러므로 y_i 에 대한 gradient를 구해 보자.
 - Gradient Descent !!

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

$$\frac{\delta C}{\delta y_i} = 4 \sum_j (p_{ij} - q_{ij})(y_i - y_j)$$



t-SNE Training

