Dimensionality Reduction techniques solve optimization problems.
Three approaches for Dimensionality Reduction:

- Distance preservation
- Topology preservation
- Information preservation

t-SNE converts Euclidean distance to similarity, which can be interpreted as probability. t-SNE is distance-based but tends to reserve topology.

1. SNE

Problem: $X = \left\{x_1,\ x_2,\ \ldots,\ x_n\ \in R^h\right\} \rightarrow Y = \left\{y_1,\ y_2,\ \ldots,\ y_n\ \in R^i\right\}$

$$\min_{Y} C(X,\ Y)$$

We have $p_{j|i}$ as the probability of x_j lies near x_i and $q_{j|i}$ as the probability of y_j lies near y_i

$$p_{j|i} = \frac{exp\left(-||x_j - x_i||^2 / 2\sigma_i^2\right)}{\sum_{k\neq i} exp\left(-||x_k - x_i||^2 / 2\sigma_i^2\right)}$$

$$q_{j|i} = \frac{exp\left(-||y_j - y_i||^2\right)}{\sum_{k\neq i} exp\left(-||y_k - y_i||^2\right)}$$

$$\Rightarrow p_{j|i} \text{ and } q_{j|i} \text{ should be the same}$$

Kullback-Leiber Divergence measures the faithfulness with which $q_{j|i}$ models $p_{j|i}$.
$P_i = \{p_{1|i},\ p_{2|i},\ \ldots,\ p_{n|i}\}$ *and* $Q_i = \{q_{1|i},\ q_{2|i},\ \ldots,\ q_{n|i}\}$ are the distributions on the neighbour of data i.

$$D_{KL}(P_i||Q_i) = \sum P(x_i) log\left(\frac{P(x_i)}{Q(x_i)}\right)$$

Loss function is defined as

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

- KL divergence is asymmetric
- KL divergence is always positive

$\Rightarrow$ Minimize loss function to find $\{y_1,\ y_2,\ \ldots,\ y_n\}$

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

$$= \sum_i \sum_j p_{j|i} log(p_{j|i}) - \sum_i \sum_j p_{j|i} log(q_{j|i})$$

$$\frac{\partial C}{\partial y_i} = \frac{\partial \sum_i \sum_j p_{j|i} log(q_{j|i})}{\partial y_i}$$

$$= \frac{\partial \left(\sum_j p_{j|i} log(q_{j|i}) + \sum_i p_{i|j} log(q_{i|j}) \right)}{\partial y_i} \quad (1)$$

$$\frac{\partial \sum_i p_{i|j} log(q_{i|j})}{\partial y_i} = \frac{\partial \sum_i p_{i|j} log(q_{i|j})}{\partial y_i} = \frac{\partial \sum_i p_{i|j} \left(\left(-\|y_j - y_i\|^2\right) - \sum_{k \neq i} \left(-\|y_k - y_j\|^2\right) \right)}{\partial y_i} \quad (2)$$

$$= 2 \sum_j p_{j|i} (y_i - y_j)$$

$$\frac{\partial \sum_i p_{j|i} log(q_{j|i})}{\partial y_i} = \frac{\partial \sum_j p_{j|i} \left(\left(-\|y_j - y_i\|^2\right) - \sum_{k \neq j} \left(-\|y_k - y_j\|^2\right) \right)}{\partial y_i} \quad (3)$$

$$= 2 \sum_j p_{j|i} \left((y_i - y_j) - \sum_{k \neq j} (y_k - y_j) \right) \quad (4)$$

From (1), (2), (4)

$$\frac{\partial C}{\partial y_i} = \frac{\partial \sum_j p_{i|j} log(q_{i|j})}{\partial y_i} + \frac{\partial \sum_i p_{i|j} log(q_{i|j})}{\partial y_i}$$

$$= 2 \sum_j p_{j|i} (y_i - y_j) + 2 \sum_i p_{i|j} \left((y_i - y_j) - \sum_{k \neq j} (y_k - y_i) \right)$$

$$= 2 \sum_j (y_i - y_j) \left(p_{j|i} + p_{i|j} - \sum_{k \neq j} (y_k - y_j) \right)$$

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