Problem 1.

For Sne:

- Given $x^1, \dots x^n \in R^D$, we define the distribution P_{ij}

$$P_{ij} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2 / 2\sigma_i^2)}$$

Find good embedding $y_1, y_2, \dots y_N \in R^d$ for $d < D$

$$Q_{ij} = \frac{\exp(-\|y_i - y_j\|^2)}{\sum_k \sum_{k \neq l} \exp(-\|y_l - y_k\|^2)} = \frac{E}{Z}$$

Optimize Q to be close to P: Minimize KL-divergence to find the embedding $y_1, \dots y_n \in R^d$:

$$KL(P\|Q) = \sum_{ij} P_{ij} \log\left(\frac{P_{ij}}{Q_{ij}}\right)$$

$$= \sum_{ij} P_{ij} \log(P_{ij}) - P_{ij} \log(Q_{ij}) \quad (1)$$

Where P_{ij} can be inferred from the data, we treat this as constant so (1) can rewritten as:

$$KL(P\|Q) = \sum_{ij} P_{ij} \log \frac{P_{ij}}{Q_{ij}}$$

$$= -\sum_{ij} P_{ij} \log Q_{ij} + \text{const}$$

$$\frac{\partial L}{\partial y_i} = \left(\frac{\partial L}{\partial d_{ij}} + \frac{\partial L}{\partial d_{ij}}\right) \frac{\partial d_{ij}}{\partial y_i} = 2 \frac{\partial L}{\partial d_{ij}} \frac{\partial d_{ij}}{\partial y_i}$$

$$\frac{\partial L}{\partial y_i} = -2 \sum P_{ij}\, \partial \log(E) - P_{ij}\, \partial \log(Z)$$

$$\partial E = -2(y_i - y_j) E$$

For T-SNE, Q_{ij} is defined as:

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

We use the same P_{ij} and our loss function is:

$$KL(P\|Q) = -\sum P_{ij}\log(E^{-1}) - P_{ij}\log(Z^{-1}) + \text{const}$$

To make derivation ~~a~~ less ~~e~~ clustered, omitting the ∂ term at the denominator.

$$\frac{\partial L}{\partial y_i} = \sum_{k, l \neq k} -P_{lk}\, \partial \log E_{kl}^{-1} + \sum_{k, l \neq k} P_{lk}\, \partial \log Z$$

First term:

$$\sum_{ij} P_{ij}\, \partial \log(E^{-1}) = -\sum_{ij} P_{ij}\, 2(y_i - y_j)\frac{E^{-2}}{E^{-1}}$$

$$= -2\sum_{ij} P_{ij}(y_i - y_j)E^{-1}$$

Second term:

$$\sum_{ij} P_{ij}\, \partial \log(Z^{-1}) = -2\sum_{ij}(y_i - y_j)\frac{1}{Z^{-1}}E^{-2}$$

$$= -2\sum_{ij}(y_i - y_j)\frac{E^{-1}}{Z^{-1}}E^{-1}$$

$$= -2\sum_{ij}(y_i - y_j)Q_{ij}E^{-1}$$

Final result:

$$\frac{\partial L}{\partial y_i} = 4\sum_{ij} P_{ij}(y_i - y_j)E^{-1} - (y_i - y_j)Q_{ij}E^{-1}$$

$$= 4\sum_{ij}(P_{ij} - Q_{ij})(y_i - y_j)E^{-1}$$

$$= 4\sum_{ij}(P_{ij} - Q_{ij})(y_i - y_j)(1+\|y_i - y_j\|^2)^{-1}$$

$$\Rightarrow \sum_{j\neq i} \cancel{-p_{j|i}} \frac{E_{ij}}{E}$$

$$\Rightarrow \sum_{ij} P_{ij}\, \partial \log(E) = \sum P_{ij}\left(-2(y_i - y_j)E\right)\frac{1}{E}$$

$$= -2 \sum_{ij} P_{ij}(y_i - y_j)$$

in the second term, $\sum_{k\neq l} P_{ij} = 1$, the derrivative is non-zero when $\forall j = i$, $k = i$ or $l = i$:

$$P_{ij}\, \partial \log(Z) = \sum \frac{1}{Z}\partial E$$

$$= 2 \sum \frac{E}{Z}(y_i - y_j)$$

$$= 2 \sum_{i\neq j} Q_{ij}(y_i - y_j)$$

$$\Rightarrow \frac{\partial L}{\partial y_i} = -2 \sum_{ij} -2P_{ij}(y_i - y_j) + 2Q_{ij}(y_i - y$$

$$= 2 \sum_{ij} 2P_{ij}(y_i - y_j) - 2Q_{ij}(y_i -$$

$$= 4 \sum_{ij} (P_{ij} - Q_{ij})(y_i - y_j)$$

PCA	t_SNE
It is a linear Dimensionality reduction technique.	It is a non-linear Dimensionality reduction technique.
It tries to preserve the global structure of the data.	It tries to preserve the local structure(cluster) of data.
It gets highly affected by outliers.	It can handle outliers
Deterministic	Randomised
Rotating vectors for preserving variance	Minimising the distance between the point in a gaussian.
Preserve using eigenvalues	Using hyper parametes