HOMEWORK 2: T-NSE

1. Biến đổi công thức toán SNE, t-NSE, có tính đạo hàm loss với các parameter.

- t-SNE is an alternative dimensionality reduction algorithm

- t-SNE tries to perserve local structure

• Low dimensional neighborhood should be the same as original neighborhood

a) Find low dimensional points such that their neighborhood distribution is similar.
The probability that point x_i chooses x_j as it neighbor:

$$p_{j|i} = \frac{e^{-\frac{\left\|x^{(i)}-x^{(j)}\right\|^2}{2\sigma_i^2}}}{\sum_{k\neq i} e^{-\frac{\left\|x^{(i)}-x^{(k)}\right\|^2}{2\sigma_i^2}}}$$

With $p_{j|i} = 0$

$$p_{j|i} = \frac{e^{-\left\|y_i-y_j\right\|^2}}{\sum_{k\neq i} e^{-\left\|y_i-y_k\right\|^2}}$$

So we now need to minimize the KL divergence values. Hence our cost function is based on this

$$C = \sum_i KL(P_i || Q_1)$$

Final distribution over pairs is symmetrized:

$$p_{ij} = \frac{1}{2N}\left(p_{i|j} + p_{j|i}\right)$$

- The parameter σ_i sets the size of the neighborhood
  - Very low σ_i - all the probability is in the nearest neighbor
  - Very high σ_i - Uniform weights
- Here we set σ_i differently for each data point

- For each distribution $P_{j|i}$(depends on σ_i) we define the perplexity

$$perp\left(P_{j|i}\right) = 2^{H\left(P_{j|i}\right)}$$

Where $H(P) = -\sum_i P_i \, log(P_i)$ is the entropy

b) Tính đạo hàm loss với các parameter.

❖ SNE (Stochastic Neighbor Embedding).

- Given $x^{(1)}, x^{(2)}, x^{(3)}, \dots, x^{(N)} \in \mathbb{R}^D$ we define te distribution P_{ij}
- We will have good embedding $y^{(1)}, y^{(2)}, y^{(3)}, \dots, y^{(N)} \in \mathbb{R}^d$ for some d < D and we can define distribution Q similary the same (notice no σ_i and not symmetric):

$$Q_{ij} = \frac{e^{-\left\|y^{(i)}-y^{(j)}\right\|^2}}{\sum_k \sum_{l \neq k} e^{-\left\|y^{(i)}-y^{(k)}\right\|^2}}$$

- Optimize Q to be close to P by use Minimize KL- divergence: **L(Q) = KL(P||Q)**

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

$$=> \frac{\partial L}{\partial y^{(i)}} = \sum_j (P_{ij} - Q_{ij})(y^{(i)} - y^{(j)})$$

❖ t-SNE (t-distribution Stochastic Neighbor Embedding).

- Probability goes to zero much slower then a Gaussian
- We have redefine Q_{ij}:

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

- We need minimize KL(P||Q):

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

$$=> \frac{\partial L}{\partial y^{(i)}} = \sum_j (P_{ij} - Q_{ij})(y^{(i)} - y^{(j)}) \left(1 + \left\|y^{(i)} - y^{(j)}\right\|^2\right)^{-1}$$

2. So sanh t-SNE and PCA

PCA	t-SNE
- PCA is a linear dimensionality reduction technique for very high dimensional data - PCA tries to find a global structure (+ Low dimensional subspace + Can lead to local inconsistencies -> Far away point can become nearest neighbors	- It is an alternative (non-linear) dimensionality reduction algorithm - It tries to perserve local structure (Low dimensional neighborhood should be the same as original neighborhood)