Derivation of SNE Gradient

Zein
GitHub@zein0115

July 16, 2020

1 Brief Review[1]

For a given high-dimensional dataset $X = \{x_1, x_2, \ldots, x_n\}$, $x_i \in \mathbb{R}^D$, define

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

For a randomized reduced dataset $Y = \{y_1, y_2, \ldots, y_n\}$, $y_i \in \mathbb{R}^M$, $M < D$, in SNE, define

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

In t-SNE, q_{ij} is defined as

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

Define the cost function in SNE

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

In t-SNE, the cost function is

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

Where

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N}$$

Using gradient descent to optimize Y, in each iteration, apply

$$y_i^{(t+1)} = y_i^{(t)} - \eta \frac{\partial C}{\partial y_i}$$

Which η is learning rate.

2 t-SNE

It might be a little complicated to find $\partial C/\partial y_i$ directly, so I define two auxiliary variables d_{ij} and Z here. [1]

$$d_{kl} = \|y_k - y_l\|, Z = \sum_{k \neq l} (1 + \|y_l - y_k\|^2)^{-1}$$

The basic idea to find the gradient is using the chain rule

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial x}$$

To simplify the derivation, we regarding C as the function of d_{ij}, d_{ij} as the function of y_i. Then, the gradient for i^{th} variable is

$$\frac{\partial C}{\partial y_i} = \sum_{k \neq l} \frac{\partial C}{\partial d_{kl}}\frac{\partial d_{kl}}{\partial y_i}$$

It could be easily derived that only $\partial d_{ij}/\partial y_i$ and $\partial d_{ji}/\partial y_i$, $i \neq j$, are nonzero, thus

$$\frac{\partial C}{\partial y_i} = \sum_j \frac{\partial C}{\partial d_{ij}}\frac{\partial d_{ij}}{\partial y_i} + \frac{\partial C}{\partial d_{ji}}\frac{\partial d_{ji}}{\partial y_i}$$

Find $\partial d_{ij}/\partial y_i$,

$$\frac{\partial d_{ij}}{\partial y_i} = \frac{\partial \|y_i - y_j\|}{\partial y_i} = \frac{\partial \sqrt{(y_i - y_j)^{\mathrm{T}}(y_i - y_j)}}{\partial y_i} = \frac{y_i - y_j}{\sqrt{(y_i - y_j)^{\mathrm{T}}(y_i - y_j)}} = \frac{y_i - y_j}{d_{ij}}$$

(Note: according to matrix calculus, the result should be $\frac{(y_i - y_j)^{\mathrm{T}}}{d_{ij}}$. Since we are going to find gradient, I will use $\frac{y_i - y_j}{d_{ij}}$ there.)

Find $\partial C/\partial d_{ij}$

$$\frac{\partial C}{\partial d_{ij}} = \frac{\sum_k \sum_l p_{kl} \log \frac{p_{kl}}{q_{kl}}}{\partial d_{ij}} = \frac{\sum_k \sum_l p_{kl}(\log p_{kl} - \log q_{kl})}{\partial d_{ij}}$$

Since p_{ij} is a constant and $p_{ii} = 0$

$$\frac{\partial C}{\partial d_{ij}} = \frac{\partial \sum_k \sum_l -p_{kl} \log q_{kl}}{\partial d_{ij}} = \frac{\partial \sum_{k \neq l} -p_{kl} \log q_{kl}}{\partial d_{ij}}$$

Using d_{ij} and Z, q_{ij} could be expressed as

$$q_{ij} = \frac{(1 + d_{ij}^2)^{-1}}{Z}$$

Then $\partial C/\partial d_{ij}$ could be expressed by

$$\frac{\partial C}{\partial d_{ij}} = \frac{\partial \sum_{k \neq l} -p_{kl} \log(1 + d_{kl}^2)^{-1}}{\partial d_{ij}} + \frac{\partial \sum_{k \neq l} p_{kl} \log Z}{\partial d_{ij}}$$

Since C is a function of d_{ij}, $\forall (i,j) \neq (k,l)$, $\partial d_{kl}/\partial d_{ij} = 0$, the first term could be eliminate to

$$\frac{\partial \sum_{k \neq l} -p_{kl} \log(1 + d_{kl}^2)^{-1}}{\partial d_{ij}} = -p_{ij} \frac{\partial \log(1 + d_{ij}^2)^{-1}}{\partial d_{ij}} = 2p_{ij} d_{ij} (1 + d_{ij}^2)^{-1}$$

Notice that $\sum_{k\neq l} p_{kl} = 1$, the second term could be eliminate to

$$\begin{aligned}\frac{\partial \sum_{k\neq l} p_{kl} \log Z}{\partial d_{ij}} &= \sum_{k\neq l} p_{kl} \frac{1}{Z} \frac{\partial Z}{\partial d_{ij}} = \sum_{k\neq l} -2p_{kl} \frac{1}{Z} d_{ij} (1 + d_{ij}^2)^{-2} \\ &= \sum_{k\neq l} -2p_{kl} q_{ij} d_{ij} (1 + d_{ij}^2)^{-1} = -2q_{ij} d_{ij} (1 + d_{ij}^2)^{-1}\end{aligned}$$

Therefore $\partial C/\partial d_{ij}$ equals to

$$\frac{\partial C}{\partial d_{ij}} = 2p_{ij} d_{ij} (1 + d_{ij}^2)^{-1} - 2q_{ij} d_{ij} (1 + d_{ij}^2)^{-1} = 2d_{ij} (1 + d_{ij}^2)^{-1} (p_{ij} - q_{ij})$$

Then the $\partial C/\partial d_{ji}$ could also be derived using the same way

$$\frac{\partial C}{\partial d_{ji}} = 2d_{ji} (1 + d_{ji}^2)^{-1} (p_{ji} - q_{ji})$$

Combine $\partial d_{ij}/\partial y_i$, $\partial C/\partial d_{ij}$ and $\partial C/\partial d_{ij}$, since $p_{ij} = p_{ji}$ and $q_{ij} = q_{ji}$, for y_i, $d_{ij} = d_{ji}$

$$\begin{aligned}\frac{\partial C}{\partial y_i} &= \sum_j \frac{y_i - y_j}{d_{ij}} 2d_{ij} (1 + d_{ij}^2)^{-1} (p_{ij} - q_{ij}) + \sum_j \frac{y_j - y_i}{d_{ji}} 2d_{ji} (1 + d_{ji}^2)^{-1} (p_{ji} - q_{ji}) \\ &= 4 \sum_j (p_{ij} - q_{ij}) (1 + \|y_i - y_j\|^2)^{-1} (y_i - y_j)\end{aligned}$$

3 SNE

Same as t-SNE, we define some auxiliary variables at first [1]

$$d_{kl} = \|y_k - y_l\|, Z_i = \sum_{k\neq i} \exp(-\|y_i - y_k\|^2) = \sum_{k\neq i} \exp(-d_{ik}^2)$$

Using the same logic in t-SNE

$$\frac{\partial C}{\partial y_i} = \sum_{k\neq l} \frac{\partial C}{\partial d_{kl}} \frac{\partial d_{kl}}{\partial y_i}$$

It could be easily derived that only $\partial d_{ij}/\partial y_i$ and $\partial d_{ji}/\partial y_i$, $i \neq j$, are nonzero, thus

$$\frac{\partial C}{\partial y_i} = \sum_j \frac{\partial C}{\partial d_{ij}} \frac{\partial d_{ij}}{\partial y_i} + \frac{\partial C}{\partial d_{ji}} \frac{\partial d_{ji}}{\partial y_i}$$

Find $\partial d_{ij}/\partial y_i$,

$$\frac{\partial d_{ij}}{\partial y_i} = \frac{\partial \|y_i - y_j\|}{\partial y_i} = \frac{\partial \sqrt{(y_i - y_j)^{\mathrm{T}} (y_i - y_j)}}{\partial y_i} = \frac{y_i - y_j}{\sqrt{(y_i - y_j)^{\mathrm{T}} (y_i - y_j)}} = \frac{y_i - y_j}{d_{ij}}$$

Find $\partial C/\partial d_{ij}$

$$\frac{\partial C}{\partial d_{ij}} = \frac{\sum_k \sum_l p_{l|k} \log \frac{p_{l|k}}{q_{l|k}}}{\partial d_{ij}} = \frac{\sum_i \sum_j p_{j|i} (\log p_{j|i} - \log q_{j|i})}{\partial d_{ij}}$$

Since $p_{j|i}$ is a constant and $p_{i|i} = 0$

$$\frac{\partial C}{\partial d_{ij}} = \frac{\partial \sum_k \sum_l -p_{l|k} \log q_{j|i}}{\partial d_{ij}} = \frac{\partial \sum_{k \neq l} -p_{l|k} \log q_{l|k}}{\partial d_{ij}}$$

Using d_{ij} and Z_i, $q_{j|i}$ could be expressed as

$$q_{j|i} = \frac{\exp(-d_{ij}^2)}{Z_i}$$

Then $\partial C / \partial d_{ij}$ could be expressed by

$$\frac{\partial C}{\partial d_{ij}} = \frac{\partial \sum_{k \neq l} -p_{l|k} \log(\exp(-d_{kl}^2))}{\partial d_{ij}} + \frac{\partial \sum_{k \neq l} p_{l|k} \log Z_k}{\partial d_{ij}}$$

Since C is a function of d_{ij}, $\forall (i,j) \neq (k,l)$, $\partial d_{kl} / \partial d_{ij} = 0$, also $p_{j|i}$ is a constant, the first term could be eliminate to

$$\frac{\partial \sum_{k \neq l} -p_{l|k} \log(\exp(-d_{kl}^2))}{\partial d_{ij}} = \frac{\partial \sum_{k \neq l} p_{l|k} d_{kl}^2}{\partial d_{ij}} = \frac{\partial p_{j|i} d_{ij}^2}{\partial d_{ij}} = 2 p_{j|i} d_{ij}$$

Now we consider the second term. Obviously, only $\partial Z_i / \partial d_{ij}$ is nonzero term, thus

$$\begin{aligned}
\frac{\partial \sum_{k \neq l} p_{l|k} \log Z_k}{\partial d_{ij}} &= \frac{\partial \sum_{l \neq i} p_{l|i} \log Z_i}{\partial d_{ij}} \\
&= \sum_{l \neq i} p_{l|i} \frac{1}{Z_i} \frac{\partial Z_i}{\partial d_{j|i}} \\
&= \sum_{l \neq i} p_{l|i} \frac{1}{Z_i} \left[-2 d_{ij} \exp(-d_{ij}^2) \right] \\
&= -2 \sum_{l \neq i} p_{l|i} q_{j|i} d_{ij} \\
&= -2 q_{j|i} d_{ij}
\end{aligned}$$

Combine two terms

$$\frac{\partial C}{\partial d_{ij}} = 2 p_{j|i} d_{ij} - 2 q_{j|i} d_{ij} = 2 d_{ij} (p_{j|i} - q_{j|i})$$

In same way, $\partial C / \partial d_{ij}$ could be expressed by

$$\frac{\partial C}{\partial d_{ij}} = 2 d_{ji} (p_{i|j} - q_{i|j})$$

Combine $\partial d_{ij} / \partial y_i$, $\partial C / \partial d_{ij}$ and $\partial C / \partial d_{ij}$

$$\begin{aligned}
\frac{\partial C}{\partial y_i} &= \sum_j \frac{\partial C}{\partial d_{ij}} \frac{\partial d_{ij}}{\partial y_i} + \frac{\partial C}{\partial d_{ji}} \frac{\partial d_{ji}}{\partial y_i} \\
&= \sum_j 2 d_{ij} (p_{j|i} - q_{j|i}) \frac{y_i - y_j}{d_{ij}} + 2 d_{ji} (p_{i|j} - q_{i|j}) \frac{y_i - y_j}{d_{ji}} \\
&= 2 \sum_j (p_{j|i} + p_{i|j} - q_{j|i} - q_{i|j})(y_i - y_j)
\end{aligned}$$

Reference

[1] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.