5/9/23

Deep Learning

- Recap
- SNE

- Recall: Create a map or an embedding of high dim. data points in a low. dim. space such that the structure in the high dim. space is retained as much as possible in the low dim. space

- $\mathcal{X} = \{x_1, x_2, \ldots x_n\}$ is the data set consisting of high. dim. data points x_i
- $\mathcal{Y} = \{y_1, y_2, \ldots y_n\}$ is the map consisting of low dim. map points y_i

- Construct pair wise distances in the high. dim. space

$$
\begin{array}{c|cccccc}
 & 1 & 2 & \cdots & j & \cdots & n \\
\hline
1 & & & & \vdots & & \\
2 & & & & \vdots & & \\
\vdots & & & & & & \\
i & - & - & - & d(x_i, x_j) & & \\
\vdots & & & & & & \\
n & & & & & &
\end{array}
\longrightarrow
\begin{array}{c|cccccc}
 & 1 & 2 & \cdots & j & \cdots & n \\
\hline
1 & & & & \vdots & & \\
2 & & & & \vdots & & \\
\vdots & & & & & & \\
P_i \rightarrow i & - & - & - & p_{j|i} & & \\
\vdots & & & & & & \\
n & & & & & &
\end{array}
$$

Distance matrix $\longrightarrow$ Similarity matrix

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

Assume that the distances map to probabilities under a Gaussian prior

- Similarly, we define a distance matrix and a similarity matrix on the map points y_i

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

- Claim: If the structure in the data points is retained in the map points, this is reflected in the similarity of the conditional distributions.

- Let's define the cost function $C = \sum_{i=1}^{n} KL(P_i \| Q_i)$ — ①

* Reading exercise: We know the $KL(p\|q)$ is assymmetric. What is its impact on the map?

- Find $\mathcal{Y} = \{y_1, y_2 \cdots y_n\}$ such that C is minimized.
- Iterative gradient descent based solution.

$$\mathcal{Y}^{(t)} = \mathcal{Y}^{(t-1)} - \eta \nabla_{\mathcal{Y}^{(t-1)}} C + \alpha(t)\left(\mathcal{Y}^{(t-1)} - \mathcal{Y}^{(t-2)}\right) \quad -②$$

$$\frac{\partial C}{\partial y_i} = 2 \sum_j \underbrace{(p_{j|i} - q_{j|i} + p_{i|j} - q_{i|j})}\,(y_i - y_j)$$