2. 希望高维和低维之间的分布尽可能相同.

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

3. 初始化时需要设置参数困惑度 (perplexity), SNE 会通过二分搜索的方法计算最佳的初始方差.

一般困惑度选择 5-50.

4. 对目标函数求梯度

$$\frac{\delta C}{\delta y_i} = 2\sum_j (p(j|i) - q(j|i) + p(i|j) - q(i|j))(y_i - y_j)$$

5. 迭代

$$Y^{(t)} = Y^{(t-1)} + \eta \frac{\partial C}{\partial Y} + \alpha(t)(Y^{(t-1)} - Y^{(t-2)})$$

t-SNE

① $p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n}$

② $q_{ij} = \frac{(1+\|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l} (1+\|y_k - y_l\|^2)^{-1}}$ (t分布) 近的越近, 远的越远

受初值影响 (RtSNE).

PCA理解 (from Cross Validated)

$X\ (n \times p)$

$C = X^T X$ (协方差矩阵应该是 $Cov = \frac{X^T X}{n-1}$)

$C = VLV^T$ (对角化, 对称矩阵正交对角化)

$X = USV^T$ (奇异值分解)

$C = X^T X = \cancel{USV}^T VSU^T USV^T = VS^2V^T = VLV^T$

$\lambda_i = S_i^2$

主成分 (坐标): $XV = USV^T V = US$.

① $X = USV^T$, V 的列向量是特征向量, 是投影方向.

② XV 或 US 是投影后的坐标

③ 协方差矩阵 C 的特征值 $\lambda_i = S_i^2$

④ 标准化的 (除以特征值的) 分数是 U 的列向量, 从 US 为主成分坐标可看出.

⑤ VS 是 loading (载荷, 变量与主成分的相关系数)

⑥ 以上的条件是 行为样本, 列为变量. 否则 U, V 变换

⑦ PCA on correlation matrix, X 的列需要中心化+标准化.

⑧ $X_k = U_k S_k V_k^T$ (U 前 k 列, S 前 $k \times k$, V^T 前 k 行).

X_k 为包含前 k 个 PCs (主成分) 的压缩数据.

⑨ if $n > p$, $U_{n \times n}$, $V_{p \times p}$, U 的最后 $n-p$ 列随意, 对应的 S 的行为 0.