$$\begin{cases} tr(B) = \frac{1}{2m}\sum_{i=1}^{m}\sum_{j=1}^{m} dist_{ij}^2 \\ \sum_{i=1}^{m} dist_{ij}^2 = tr(B) + m b_{jj} \\ \sum_{j=1}^{m} dist_{ij}^2 = tr(B) + m b_{ii} \end{cases}$$

$b_{ij} = \frac{1}{2}(b_{ii} + b_{jj} - dist_{ij}^2)$

又由于：$B = Y^TY$，B是对称矩阵，正交对角化

$B = Q\Lambda Q^T = Y^TY$

$Y^T = Q\Lambda^{\frac{1}{2}}$

主要思想推导：

$$\begin{cases} dist_{ij} = d(x_i, x_j) = \|Y_i - Y_j\| \\ B = Y^TY, \boxed{B_{ij} = b_{ij} = d(Y_i, Y_j) = \|Y_i - \cancel{Y_j}\|} \end{cases}$$

(方便计算).

注意 ①当 dist 矩阵本身为欧氏距离 MDS = PCA

②MDS强调对不同种类 dist 都可转为欧氏距离.

t-SNE.

①高维空间概率分布 F_1

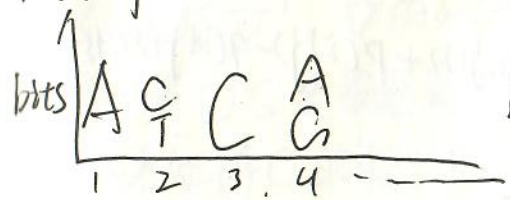
②低维空间概率分布 F_2

③使 F_1 与 F_2 尽可能相同，通过迭代.

香农信息熵

$H = -\sum_{i=1}^{n} P_i \log(P_i)$

motif



通过信息熵计算.

KL散度：衡量两个分布是否相似

$D_{KL}(P\|Q) = \sum_i P(i) \ln \frac{P(i)}{Q(i)}$

当两个概率分布完全相同时，$D_{KL} = 0$，否则 $D_{KL} > 0$.

SNE的数学过程

1. 先通过欧氏距离表示点的概率分布

高维情况：$P_{j|i} = \frac{\exp(-\|x_i - x_j\|^2/(2\sigma_i^2))}{\sum_{k\neq i}\exp(-\|x_i - x_k\|^2/(2\sigma_i^2))}$ $\frac{(以x_j为中心)}{(以入非x_i为中心)}$

低维情况：$q_{j|i} = \exp(-\|y_i - y_j\|^2)$