标题：Information-theoretic metric learning

主要介绍了马氏距离的分类方法

马氏距离定义如下：

$$d_A(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i - \boldsymbol{x}_j)^T A (\boldsymbol{x}_i - \boldsymbol{x}_j).$$

当距离为欧氏距离时，其中的 A 为单位矩阵；很显然，马氏距离是欧氏距离的推广。度量学习的核心就在于计算 A 矩阵。

设 X 为输入向量，其格拉姆矩阵为：$K_0 = X^T X.$，我们需要得到这么一个矩阵 K，其满足：

$$\begin{array}{rll} \min & D_{\text{Burg}}(K, K_0) & \\ \text{subject to} & K_{ii} + K_{jj} - 2K_{ij} \leq u & (i,j) \in S, \\ & K_{ii} + K_{jj} - 2K_{ij} \geq l & (i,j) \in D, \\ & K \succeq 0. & \end{array}$$

其中最小化目标称为 bregman 散度，定义如下：

$$D_{\text{Burg}}(K, K_0) = \text{Tr}(KK_0^{-1}) - \log\det(KK_0^{-1}) - n.$$

Bregman 散度是可以解出来的，设为 W，其格拉姆矩阵即为目标矩阵 A

详细的算法如下：

ALGORITHM 1: Algorithm for information-theoretic metric learning

ITMETRICLEARN(X, S, D, u, l)

Input: X: input $d \times n$ matrix, S: set of similar pairs, D: set of dissimilar pairs, u, l: distance thresholds

Output: W: output factor matrix, where $W^T W = A$

1. Set $W = I_d$ and $\lambda_{ij} = 0 \; \forall \, i, j$
2. Repeat until convergence:
   - Pick a constraint $(i, j) \in S$ or $(i, j) \in D$
   - Let $\boldsymbol{v}^T$ be row i of X minus row j of X
   - Set the following variables:
     1. $\boldsymbol{w} = W\boldsymbol{v}$
     2. **if** (similarity constraint)

        $\gamma = \min\left(\lambda_{ij}, \frac{1}{\|\boldsymbol{w}\|_2^2} - \frac{1}{u}\right)$

        $\beta = \gamma/(1 - \gamma\|\boldsymbol{w}\|_2^2)$

        else if (dissimilarity constraint)

        $\gamma = \min\left(\lambda_{ij}, \frac{1}{l} - \frac{1}{\|\boldsymbol{w}\|_2^2}\right)$

        $\beta = -\gamma/(1 + \gamma\|\boldsymbol{w}\|_2^2)$
     3. $\lambda_{ij} = \lambda_{ij} - \gamma$
   - Compute the Cholesky factorization $LL^T = I + \beta \boldsymbol{w}\boldsymbol{w}^T$
   - Set $W \leftarrow L^T W$
3. Return W