

实验1. 度量学习实验报告

MG1733099, 周天烁, tianshuo.zhou@smail.nju.edu.cn

2017 年 11 月 7 日

1 综述

距离度量学习 (distance metric learning) 试图在样本空间中学习到一种距离计算方法使得样本中有用的信息得以充分的利用, 同时尽可能地删减对于特定学习目标来说无效或者冗余的信息。当学习得到的距离表示比原样本简化许多的时候, 可以起到降维的作用。距离度量学习方法有很多种, 主要分为有监督度量学习 (如NCA), 无监督度量学习, 最大间隔度量学习 (如LMNN) 和基于核方法的度量学习。近邻成分分析 (Neighbourhood Component Analysis) 是一种基于马氏距离 (Mahalanobis distance) 的有监督距离度量学习。目标函数为选择同类样本作为分类近邻的概率。通过最大化目标函数得到马氏距离中的参数矩阵M, i.e. $M=A^T A$ 中的矩阵A。

2 任务1

度量函数学习目标

NCA中马氏距离的定义如下:

$$d(x, y) = (x - y)^T Q (x - y) \quad (1)$$

其中, $Q=A^T A$, 可以把A理解为原向量的线性变换函数。对于每一个样本 X^i 选择样本 X^j 作为其近邻的概率 p_{ij} 可以softmax表示如下:

$$p_{ij} = \frac{\exp(-\|Ax_i - Ax_j\|^2)}{\sum_{k \neq i} \exp(-\|Ax_i - Ax_k\|^2)} \quad (2)$$

算法的目标就是最大化样本 X^i 被正确分类的概率, 函数表达如下:

$$f(A) = \sum_i \sum_{j \in C_i} p_{ij} = \sum_i p_i \quad (3)$$

其中, p_i 表示样本被正确分类的概率:

$$p_i = \sum_{j \in C_i} p_{ij} \quad (4)$$

优化算法

由于目标函数较为简单，可以直接利用梯度下降相关方法进行优化，梯度计算公式如下：

$$\frac{\partial f}{\partial A} = 2A \sum_i \left(p_i \sum_k p_{ik} x_{ik} x_{ik}^T - \sum_{j \in C_i} p_{ij} x_{ij} x_{ij}^T \right) \quad (5)$$

本任务中样本较少，因此算法收敛需要迭代的次数较多（大于一千次）。

3 任务2

与任务一优化算法相同。由于本次实验的样本分布比较随机和均匀，因此直接采用批量梯度下降遍历所有样本一次作为一个epoch，直至收敛。

参考文献

Goldberger, J., G.E. Hinton, S.T. Roweis, and R.R. Salakhutdinov. (2005). "Neighbourhood Components Analysis." In *Advances in Neural Information Processing System* 17 (NIPS).

An Overview of Distance Metric Learning. October 28, 2007. Liu Yang.

周志华 (2016). 机器学习. 清华大学出版社, 北京.