Boosting and Clustering

zyt

ustc

2022 年 11 月 12 日

Content

8.2

对于 0/1 损失函数来说, 指数损失函数并非仅有的一致替代函数. 考虑式 (8.5), 试证明: 任意损失函数 $\ell(-f(\boldsymbol{x})H(\boldsymbol{x}))$, 若对于 $H(\boldsymbol{x})$ 在区间 $[-\infty, \delta](\delta > 0)$ 上单调递减, 则 ℓ 是 0/1 损失函数的一致替代函数.
Hint:

$$\begin{aligned} L(H \mid \mathcal{D}) &= \mathbb{E}_{\boldsymbol{x}\sim\mathcal{D}}[\ell(-yH(\boldsymbol{x}))] \\ &= P(y=1 \mid \boldsymbol{x})\ell(-H(\boldsymbol{x})) + P(y=-1 \mid \boldsymbol{x})\ell(H(\boldsymbol{x})) \end{aligned}$$

$$\frac{\partial L(H \mid \mathcal{D})}{\partial H(\boldsymbol{x})} = P(\boldsymbol{x})(-P(y=1 \mid \boldsymbol{x})\ell(-H(\boldsymbol{x})) + P(y=-1 \mid \boldsymbol{x})\ell(H(\boldsymbol{x}))) = 0$$

We can get

$$P(y=1 \mid \boldsymbol{x})\ell(-H(\boldsymbol{x})) = P(y=-1 \mid \boldsymbol{x})\ell(H(\boldsymbol{x}))$$

Suppose $P(y=1 \mid \boldsymbol{x}) > P(y=-1 \mid \boldsymbol{x})$, get $\text{sign}(H(x)) = 1$

给定任意的两个相同长度向量 x, y ，其余弦距离为 $1 - \frac{x^\top y}{|x|y|}$ ，证明余弦距离不满足传递性，而余弦夹角 $\arccos\left(\frac{x^\top y}{|x||y|}\right)$ 满足

Hint: transitivity: $d(x, y) + d(y, z) \geq d(x, z)$

For cosine distance, this equals to $(x - y)^{\mathrm{T}}(x + z) \geq 0$, which is easy to validate.

For arccosine, WLOG we can suppose x, y, z both are unit vectors, then this euqal to $\cos(\arccos x^T z) \geq \cos(\arccos x^T y + \arccos y^T z)$

$$\Leftrightarrow x^T z \geq (x^T y)(y^T z) - \sqrt{(1 - x^T y)^2 (1 - y^T z)^2}$$

Some errors: directly from geometric/ in above using $(x^T y)(y^T z) = x^T (y y^T) z$, only consider 3d situation.

kmeans 算法收敛性

Hint:

We show that the loss function is guaranteed to decrease monotonically in each iteration.

EM convergence. See PRML P425 for details.

Some errors: Only prove the corroctness of center choosen.

在k-means算法中替换欧式距离为其他任意的度量, 请问“聚类中心如何计算?

Hint: Object function: $L = \sum_{i=1}^{k} \sum_{\boldsymbol{x} \in C_i} \operatorname{dist}(\boldsymbol{x}, \boldsymbol{\mu}_i)$

If L is idfferentiable, then we let $\frac{\partial E}{\partial \boldsymbol{\mu}_i} = 0 (i = 1, \ldots, k)$

ch8 Ensemble

1. Are all ensemble make model better?
2. Boosting, why adaboost is useful? How the parameters change in each iteration in adaboost? The procedure of GBDT.
3. Bagging, why bagging is useful? The procedure of stacking.

ch9 Clustering

1. Metrics: Jaccard, FM, Rand, DBI, DI
2. Distance: Minkowski, VDM
3. Methods: Kmeans(convergence proof, choosen of initial parameters), LVQ, GMM, DBSCAN, AGNES(different linkage) and the difference of this methods(i.e. in special situation, which one is better)
4. Theoretical Analysis: convergence of kmeans, convergence of GMM,