Homework8&9

王世炟 PB20151796 2022/11/10

HW8

8.2

损失函数 l(-f(x)H(x)) = l(-H(x)) * P(f(x) = 1|x) + l(H(x)) * P(f(x) = -1|x)

为了最小化 L, 我们分情况讨论:

若
$$P(f(\boldsymbol{x})=1|\boldsymbol{x})>P(f(\boldsymbol{x})=-1|\boldsymbol{x})$$
,我们希望 $l(-H(\boldsymbol{x}))< l(H(\boldsymbol{x}))$ 由于 $l(-f(\boldsymbol{x})H(\boldsymbol{x}))$ 对于 $H(\boldsymbol{x})$ 单调递减,所以 $-H(\boldsymbol{x})>H(\boldsymbol{x})$

8.8

MultiBoosting 方法结合了 Bagging 与 AdaBoost , Bagging 可以降低 AdaBoost ;但是训练以及测试成本均会增加。

Iterative Bagging 会降低偏差,但也会使方差上升。

HW9

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 $\arccos(\frac{(x^\top y)}{|x||y|})$ 满足.

余弦距离不满足传递性:

反例:

$$m{x} = (1,0)^T$$
 $m{y} = (0,1)^T$ $m{z} = (rac{\sqrt{2}}{2},rac{\sqrt{2}}{2})^T$ $d_{ ext{\frac{\frac{1}}{2}}}(x,y) = 1$ $d_{ ext{\frac{1}{2}}}(x,z) = 1 - rac{\sqrt{2}}{2}$ $d_{ ext{\frac{1}{2}}}(z,y) = 1 - rac{\sqrt{2}}{2}$ $d_{ ext{\frac{1}{2}}}(x,z) + d_{ ext{\frac{1}{2}}}(z,y) = 2 - \sqrt{2} < 1 = d_{ ext{\frac{1}{2}}}(x,y)$ \therefore 余弦距离不满足传递性

余弦夹角满足传递性:

即证:

$$egin{aligned} & rccos(rac{(oldsymbol{x}^{ op}oldsymbol{y})}{|oldsymbol{x}||oldsymbol{y}|}) \leq rccos(rac{(oldsymbol{x}^{ op}oldsymbol{z})}{|oldsymbol{x}||oldsymbol{z}|}) + rccos(rac{(oldsymbol{z}^{ op}oldsymbol{y})}{|oldsymbol{z}||oldsymbol{y}|}) \ & ext{由于} \cos(oldsymbol{x}) + egin{aligned} & rccos(rac{(oldsymbol{z}^{ op}oldsymbol{y})}{|oldsymbol{x}||oldsymbol{y}|}) + rccos(rac{(oldsymbol{z}^{ op}oldsymbol{y})}{|oldsymbol{z}||oldsymbol{y}|})) \end{aligned}$$

由

$$\cos(\alpha + \beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta) = \cos(\alpha)\cos(\beta) - \sqrt{(1 - \cos^2(\alpha))(1 - \cos^2(\beta))}$$

所以只需证

$$rac{oldsymbol{(oldsymbol{x}^ opoldsymbol{y})}}{|oldsymbol{x}||oldsymbol{y}|} \geq rac{oldsymbol{(oldsymbol{x}^ opoldsymbol{z})}}{|oldsymbol{x}||oldsymbol{z}|} * rac{oldsymbol{(oldsymbol{z}^ opoldsymbol{y})}}{|oldsymbol{z}||oldsymbol{y}|} - \sqrt{ig(1-(rac{oldsymbol{(oldsymbol{x}^ opoldsymbol{z})}}{|oldsymbol{x}||oldsymbol{z}|}ig)^2ig)} * ig(1-(rac{oldsymbol{(oldsymbol{z}^ opoldsymbol{y})}}{|oldsymbol{z}||oldsymbol{y}|}ig)^2ig)$$

而上式的右侧

$$=\frac{\boldsymbol{x}^{\top}(\boldsymbol{z}\boldsymbol{z}^{\top})\boldsymbol{y}}{|\boldsymbol{x}||\boldsymbol{z}||\boldsymbol{z}||\boldsymbol{y}|} - \sqrt{(1 - (\frac{(\boldsymbol{x}^{\top}\boldsymbol{z})}{|\boldsymbol{x}||\boldsymbol{z}|})^2) * (1 - (\frac{(\boldsymbol{z}^{\top}\boldsymbol{y})}{|\boldsymbol{z}||\boldsymbol{y}|})^2)}$$

$$=\frac{(\boldsymbol{x}^{\top}\boldsymbol{y})}{|\boldsymbol{x}||\boldsymbol{y}|} - \sqrt{(1 - (\frac{(\boldsymbol{x}^{\top}\boldsymbol{z})}{|\boldsymbol{x}||\boldsymbol{z}|})^2) * (1 - (\frac{(\boldsymbol{z}^{\top}\boldsymbol{y})}{|\boldsymbol{z}||\boldsymbol{y}|})^2)} \le \frac{(\boldsymbol{x}^{\top}\boldsymbol{y})}{|\boldsymbol{x}||\boldsymbol{y}|}$$

得证

2.

证明k-means算法的收敛性。

即证明损失函数 $E=\sum_{i=1}^k\sum_{m{x}\in C_i}||m{x}-m{\mu}_i||_2^2$ 在算法执行过程中单调递减且有界。

单调性:

首先,算法分为两步:

- 1. 就近分配 x_i 的簇
- 2. 更新均值向量

对于第一步,如果 x_i 被分配的簇不变,则损失函数不变,即不增

如果 x_i 被分配的簇改变, 考虑以下式子:

$$E' = E - ||\boldsymbol{x}_i - \boldsymbol{\mu}_k||_2^2 + ||\boldsymbol{x}_i - \boldsymbol{\mu}_{k'}||_2^2$$
 因为 $||\boldsymbol{x}_i - \boldsymbol{\mu}_k||_2^2 > ||\boldsymbol{x}_i - \boldsymbol{\mu}_{k'}||_2^2$ ∴ 损失函数减小

所以对于第一步损失函数不增;

对于第二步,考虑一个簇的损失函数:

$$E = \sum_{oldsymbol{x} \in C_i} ||oldsymbol{x} - oldsymbol{\mu}_i||_2^2$$

令 $\frac{\partial E}{\partial \mu_i} = 0$ 可得:

$$oldsymbol{\mu}_i = rac{1}{|C_i|} \sum_{oldsymbol{x} \in C_i} oldsymbol{x} = oldsymbol{\mu}_i'$$

即更新之后的 μ'_i 是使损失函数最小的点,也不会导致损失函数增加。

故损失函数在算法执行过程中单调递减

损失函数有下界 0

故k-means算法收敛。

3.

在k-means算法中替换欧氏距离为其他任意的度量,请问聚类簇"中心"如何计算?

若损失函数是凸的,则直接对其求梯度并令其等于0,可得到聚类中心。即用到一个簇中所有点距离均值最小的点作为簇的"中心"。