HW9

PB20111689 蓝俊玮

T1

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

取反例 $A=(1,0),B=(1,1),C=(0,1)$，则可以计算得到
$dist(A,B)=1-\frac{\sqrt{2}}{2}$，$dist(B,C)=1-\frac{\sqrt{2}}{2}$，$dist(A,C)=1$，得到
$2-\sqrt{2}=dist(A,B)+dist(B,C)<dist(A,C)=1$，所以余弦距离在某些情况下是不满足传递性的。

而对于余弦夹角来说，要证明 $\arccos\left(\frac{\boldsymbol{x}^T\boldsymbol{y}}{|\boldsymbol{x}||\boldsymbol{y}|}\right)\le\arccos\left(\frac{\boldsymbol{x}^T\boldsymbol{z}}{|\boldsymbol{x}||\boldsymbol{z}|}\right)+\arccos\left(\frac{\boldsymbol{z}^T\boldsymbol{y}}{|\boldsymbol{z}||\boldsymbol{y}|}\right)$。

由 $\cos(\alpha+\beta)=\cos\alpha\cos\beta-\sin\alpha\sin\beta=\cos\alpha\cos\beta-\sqrt{1-\cos^2\alpha}\sqrt{1-\cos^2\beta}$ 可以得到：
$\alpha+\beta=\arccos\left(\cos\alpha\cos\beta-\sqrt{1-\cos^2\alpha}\sqrt{1-\cos^2\beta}\right)$

则等价证明为：
$\arccos\left(\frac{\boldsymbol{x}^T\boldsymbol{y}}{|\boldsymbol{x}||\boldsymbol{y}|}\right)\le\arccos\left(\frac{\boldsymbol{x}^T\boldsymbol{z}}{|\boldsymbol{x}||\boldsymbol{z}|}\right)+\arccos\left(\frac{\boldsymbol{z}^T\boldsymbol{y}}{|\boldsymbol{z}||\boldsymbol{y}|}\right)=\arccos\left(\frac{\boldsymbol{x}^T\boldsymbol{z}}{|\boldsymbol{x}||\boldsymbol{z}|}\frac{\boldsymbol{z}^T\boldsymbol{y}}{|\boldsymbol{z}||\boldsymbol{y}|}-\sqrt{1-(\frac{\boldsymbol{x}^T\boldsymbol{z}}{|\boldsymbol{x}||\boldsymbol{z}|})^2}\sqrt{1-(\frac{\boldsymbol{z}^T\boldsymbol{y}}{|\boldsymbol{z}||\boldsymbol{y}|})^2}\right)$
即证明：$\arccos\left(\frac{\boldsymbol{x}^T\boldsymbol{y}}{|\boldsymbol{x}||\boldsymbol{y}|}\right)\le\arccos\left(\frac{\boldsymbol{x}^T\boldsymbol{z}}{|\boldsymbol{x}||\boldsymbol{z}|}\frac{\boldsymbol{z}^T\boldsymbol{y}}{|\boldsymbol{z}||\boldsymbol{y}|}-\sqrt{1-(\frac{\boldsymbol{x}^T\boldsymbol{z}}{|\boldsymbol{x}||\boldsymbol{z}|})^2}\sqrt{1-(\frac{\boldsymbol{z}^T\boldsymbol{y}}{|\boldsymbol{z}||\boldsymbol{y}|})^2}\right)$。由于 $\arccos$ 函数在 $[-1,1]$ 上是单调递减的，则需证明：$\frac{\boldsymbol{x}^T\boldsymbol{y}}{|\boldsymbol{x}||\boldsymbol{y}|}\ge\frac{\boldsymbol{x}^T\boldsymbol{z}}{|\boldsymbol{x}||\boldsymbol{z}|}\frac{\boldsymbol{z}^T\boldsymbol{y}}{|\boldsymbol{z}||\boldsymbol{y}|}-\sqrt{1-(\frac{\boldsymbol{x}^T\boldsymbol{z}}{|\boldsymbol{x}||\boldsymbol{z}|})^2}\sqrt{1-(\frac{\boldsymbol{z}^T\boldsymbol{y}}{|\boldsymbol{z}||\boldsymbol{y}|})^2}$，两边同时乘 $|\boldsymbol{x}||\boldsymbol{y}||\boldsymbol{z}|^2$，即需要证明：$\sqrt{|\boldsymbol{x}|^2|\boldsymbol{z}|^2-(\boldsymbol{x}^T\boldsymbol{z})^2}\sqrt{|\boldsymbol{z}|^2|\boldsymbol{y}|^2-(\boldsymbol{z}^T\boldsymbol{y})^2}\ge\boldsymbol{x}^T\boldsymbol{z}\boldsymbol{z}^T\boldsymbol{y}-|\boldsymbol{z}|^2\boldsymbol{x}^T\boldsymbol{y}$，两边平方得到：$(|\boldsymbol{x}|^2|\boldsymbol{z}|^2-(\boldsymbol{x}^T\boldsymbol{z})^2)(|\boldsymbol{z}|^2|\boldsymbol{y}|^2-(\boldsymbol{z}^T\boldsymbol{y})^2)\ge(\boldsymbol{x}^T\boldsymbol{z}\boldsymbol{z}^T\boldsymbol{y}-|\boldsymbol{z}|^2\boldsymbol{x}^T\boldsymbol{y})^2$。即需要证明这个不等式成立。

设 $\boldsymbol{x}=(x_1,x_2,\dots,x_n)^T,\boldsymbol{y}=(y_1,y_2,\dots,y_n)^T,\boldsymbol{z}=(z_1,z_2,\dots,z_n)^T$

$$
\begin{aligned}
|\boldsymbol{x}|^2|\boldsymbol{z}|^2-(\boldsymbol{x}^T\boldsymbol{z})^2&=(x_1^2+x_2^2+\dots+x_n^2)(z_1^2+z_2^2+\dots+z_n^2)-(x_1z_1+x_2z_2+\dots+x_nz_n)^2\\
&=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_iz_j-x_jz_i)^2
\end{aligned}
$$

$$
|\boldsymbol{z}|^2|\boldsymbol{y}|^2-(\boldsymbol{z}^T\boldsymbol{y})^2=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(z_iy_j-z_jy_i)^2
$$

$$
\begin{aligned}
(\boldsymbol{x}^T\boldsymbol{z}\boldsymbol{z}^T\boldsymbol{y}-|\boldsymbol{z}|^2\boldsymbol{x}^T\boldsymbol{y})^2&=((x_1z_1+\dots+x_nz_n)(z_1y_1+\dots+z_ny_n)-(z_1^2+\dots+z_n^2)(x_1y_1+\dots+x_ny_n))^2\\
&=(\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_iz_j-x_jz_i)(z_iy_j-z_jy_i))^2
\end{aligned}
$$

则由柯西不等式

$$
(\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_iz_j-x_jz_i)^2)(\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(z_iy_j-z_jy_i)^2)\ge(\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_iz_j-x_jz_i)(z_iy_j-z_jy_i))^2
$$

即得到证明：

$$
(|\boldsymbol{x}|^2|\boldsymbol{z}|^2-(\boldsymbol{x}^T\boldsymbol{z})^2)(|\boldsymbol{z}|^2|\boldsymbol{y}|^2-(\boldsymbol{z}^T\boldsymbol{y})^2)\ge(\boldsymbol{x}^T\boldsymbol{z}\boldsymbol{z}^T\boldsymbol{y}-|\boldsymbol{z}|^2\boldsymbol{x}^T\boldsymbol{y})^2
$$

因此证明出余弦夹角满足传递性。

T2

> 证明 k-means 算法的收敛性

k-means 的损失函数为 $E=\sum_{i=1}^{k}\sum_{\boldsymbol{x}\in C_i}||\boldsymbol{x}-\boldsymbol{\mu}_i||_2^2$，则有 $\frac{\partial E}{\partial \boldsymbol{\mu}_i}=2\sum_{\boldsymbol{x}\in C_i}(\boldsymbol{x}-\boldsymbol{\mu}_i)=0$ 得到 $\boldsymbol{\mu}_k=\frac{1}{|C_i|}\sum_{\boldsymbol{x}\in C_i}\boldsymbol{x}=\boldsymbol{\mu}_k'$，可以得知在更新之后的均值向量 $\boldsymbol{\mu}_k'$ 是损失函数最小值的一个极值点。那么就说明了，在每次更新中心点为均值向量时，都能让损失函数 E 变得更小。因此 k-means 算法的更新能够让损失函数 E 单调递减，同时又因为 $E\geq 0$ 是有界的，因此 k-means 算法具有收敛性。

T3

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

当不再采用欧式距离时，则可以通过计算每个聚类簇中的距离度量之和最小的点作为中心点。即对于一个簇 C_i，选取 $x_0=\arg\min_{x_0\in C_i}\sum_{x\in C_i}dist(x_0,x)$，其中 $dist(x_0,x)$ 为新的度量方式。即从一个聚类簇中，选取这样一个点：它到这个聚类簇中其它的所有点的距离度量之和最小，并将这个点作为该聚类簇的中心点。