第三章 度量与投影

第 6 讲 内积与范数：数据度量的观点

黄定江

DaSE @ ECNU
djhuang@dase.ecnu.edu.cn

内积和范数引例：文本聚类

在第 2.1.1 节中，我们对纽约时报在 2010 年 12 月 7 日的四则新闻提要都进行了向量化的表示，我们希望知道这四则新闻提要有哪些是相关的，可以通过对这四则新闻提要进行简单聚类来实现:

(a) Suit Over Targeted Killing in Terror Case Is Dismissed $\cdots$
(b) In Tax Deal With G.O.P, a Portent for the Next 2 Years $\cdots$
(c) Obama Urges China to Check North Koreans $\cdots$
(d) Top Test Scores From Shanghai Stun Educators $\cdots$

$$sim_{cos}(\boldsymbol{x}, \boldsymbol{y}) = \frac{\boldsymbol{x} \cdot \boldsymbol{y}}{|\boldsymbol{x}||\boldsymbol{y}|},$$

其中 $\boldsymbol{x}, \boldsymbol{y}$ 是文本向量，$sim_{cos}(\boldsymbol{x}, \boldsymbol{y})$ 表示 $\boldsymbol{x}, \boldsymbol{y}$ 的余弦相似度。

内积和范数引例：手写数字分类

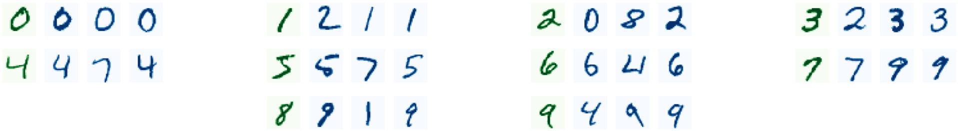

图 1: 对 MINST 数据集进行分类（绿色的为训练集，蓝色的为测试集）

$$d(A,B)=\sum_{jk}|\boldsymbol{A}_{jk}-\boldsymbol{T}_{jk}|$$

其中 d 是手写数字训练图片的表示矩阵 $\boldsymbol{A}$ 和测试图片的表示矩阵 $\boldsymbol{T}$ 之间的距离（两个矩阵同等大小），j,k 取遍矩阵所有元素。

6.1.1 向量范数：复数的模

例 1

复数 $\boldsymbol{x} = (a, b) = a + ib$ 的长度或者模指的是

$$||\boldsymbol{x}|| = \sqrt{a^2 + b^2}$$

显然复向量 $\boldsymbol{x}$ 的模 $||\boldsymbol{x}||$ 具有下列三条性质：

(1) $||\boldsymbol{x}|| \geq 0$，当且仅当 $\boldsymbol{x} = 0$ 时等号成立；

(2) $||\lambda \boldsymbol{x}|| = |\lambda| ||\boldsymbol{x}||; (\forall \lambda \in \mathbb{R})$

(3) $||\boldsymbol{x} + \boldsymbol{y}|| \leq ||\boldsymbol{x}|| + ||\boldsymbol{y}||. (\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C})$

向量的模

例 2

n 维向量 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)\in\mathbb{R}^n$ 的模或长度定义为

$$||\boldsymbol{x}||=\sqrt{x_1^2+x_2^2+\cdots+x_n^2}$$

显然向量 $\boldsymbol{x}$ 的模 $||\boldsymbol{x}||$ 也具有下列三条性质：

(1) $||\boldsymbol{x}||\geq 0$，当且仅当 $\boldsymbol{x}=0$ 时等号成立；

(2) $||\lambda\boldsymbol{x}||=|\lambda|||\boldsymbol{x}||;(\forall\lambda\in\mathbb{R})$

(3) $||\boldsymbol{x}+\boldsymbol{y}||\leq||\boldsymbol{x}||+||\boldsymbol{y}||.(\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n)$

向量范数的定义

定义 1

设 $\mathbb{V}$ 是数域上 $\mathbb{K}$ 的 n 维线性空间，函数

$$\begin{aligned} \|\cdot\|: \mathbb{V} &\to \mathbb{R}, \\ \boldsymbol{x} &\mapsto \|\boldsymbol{x}\|, \end{aligned}$$

它把向量 $\boldsymbol{x}$ 映射为它的长度 $\|\boldsymbol{x}\| \in \mathbb{R}$，并且使得对 $\forall \lambda \in \mathbb{R}$ 和 $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{V}$，满足

(1) 非负性：$\|\boldsymbol{x}\| \geq 0$，$\|\boldsymbol{x}\| = 0$ 当且仅当 $\boldsymbol{x} = 0$；

(2) 齐次性：$\|\lambda \boldsymbol{x}\| = |\lambda| \, \|\boldsymbol{x}\|$；

(3) 三角不等式：$\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$；

称 $\|\boldsymbol{x}\|$ 是向量 $\boldsymbol{x}$ 的**向量范数**，称定义了范数的线性空间 $\mathbb{V}$ 为**赋范线性空间**。

例 3

对任给的 $\boldsymbol{x}=(x_1,x_2,x_3)^{\mathrm{T}}\in\mathbb{C}^3$，试问如下实值函数是否构成向量范数？

1. $|x_1|+|2x_2+x_3|$,
2. $|x_1|+|2x_2|-5|x_3|$.

解

1. 非负性：$|x_1|+|2x_2+x_3|\geq 0$;
   齐次性：令 $c\in\mathbb{C}, |cx_1|+|2cx_2+cx_3|=|c|(|x_1|+|2x_2+x_3|)$;
   三角不等式：令 $\boldsymbol{x}=(x_1,x_2,x_3)^{\mathrm{T}}, \boldsymbol{y}=(y_1,y_2,y_3)^{\mathrm{T}}\in\mathbb{C}^3$，则
   $|x_1+y_1|+|2(x_2+y_2)+(x_3+y_3)|\leq|x_1|+|2x_2+x_3|+|y_1|+|2y_2+y_3|$.
2. 取 $\boldsymbol{x}=(0,0,1)$ 则 $|0|+|2\times 0|-5|1|=-5<0$ 不满足非负性。

6.1.2 几种常用的向量范数：l_p 范数

例 4

对于任意 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)\in\mathbb{R}^n$，由

$$\|\boldsymbol{x}\|_p=(\sum_{i=1}^{n}|x_i|^p)^{\frac{1}{p}},1\leq p<\infty$$

定义的 $||\cdot||_p$ 是 $\mathbb{R}^n$ 上的向量范数，称为p **范数**或 l_p **范数**。

(1) 当 $p=1$ 时，得到 **1 范数**或 l_1 范数，也称为 Manhattan 范数

$$\|\boldsymbol{x}\|_1=\sum_{i=1}^{n}|x_i|$$

(2) 当 $p=2$ 时，得到 **2 范数**或 l_2 范数，也称为欧几里得范数

$$\|\boldsymbol{x}\|_2=\sqrt{\sum_{i=1}^{n}x_i^2}$$

l_∞ 范数

在例4中，在广义实数范围内，p 能否取到正无穷大呢？具体而言，如何计算这种范数呢？

例 5

对于任意 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)\in\mathbb{R}^n$，由

$$\|\boldsymbol{x}\|_\infty=\lim_{p\to\infty}\|\boldsymbol{x}\|_p,$$

也就是，

$$\|\boldsymbol{x}\|_\infty=\max_{i=1,\cdots,n}|x_i|,$$

定义的 $||\cdot||_\infty$ 是 $\mathbb{R}^n$ 上的向量范数，称为∞ **范数**或 l_∞ 范数。

l_∞ 范数

证明.

验证 $\|\boldsymbol{x}\|_\infty \equiv \max_i |\boldsymbol{x}_i|$ 是向量范数显然很容易。下证 $\max_i |\boldsymbol{x}_i| = \lim_{p\to+\infty} \|\boldsymbol{x}\|_p$。令 $\|\boldsymbol{x}_j\| = \max_j |\boldsymbol{x}_i|$，则有

$$\begin{aligned}\|\boldsymbol{x}\|_\infty = |\boldsymbol{x}_j| &\le (\sum_{i=1}^{n} |\boldsymbol{x}_i|^p)^{(1/p)} = \|\boldsymbol{x}\|_p \\ &\le (n|\boldsymbol{x}_j|^p)^{(1/p)} = n^{(1/p)}\|\boldsymbol{x}\|_\infty\end{aligned}$$

由极限的夹逼准则，并注意到 $\lim_{p\to+\infty} n^{1/p} = 1$，即得欲证结论。 □

非向量范数

例 6

当 $0 < p < 1$，由

$$\|\boldsymbol{x}\|_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$$

定义的 $\|\cdot\|_p$ 不是 $\mathbb{R}^n$ 上的向量范数。

证明.

考虑 $n = 2, p = \frac{1}{2}$. 取 $\boldsymbol{\alpha} = (1,0)^{\mathrm{T}}, \boldsymbol{\beta} = (0,1)^{\mathrm{T}}$, 则

$$\|\boldsymbol{\alpha}\|_{\frac{1}{2}} = \|\boldsymbol{\beta}\|_{\frac{1}{2}} = 1, \|\boldsymbol{\alpha} + \boldsymbol{\beta}\|_{\frac{1}{2}} = 4$$

$$\|\boldsymbol{\alpha} + \boldsymbol{\beta}\|_{\frac{1}{2}} \geq \|\boldsymbol{\alpha}\|_{\frac{1}{2}} + \|\boldsymbol{\beta}\|_{\frac{1}{2}}$$

□

基数函数：l_0 范数

定义 2

向量 $\boldsymbol{x}$ 的**基数函数**定义为 $\boldsymbol{x}$ 中非零元素的个数，即

$$card(\boldsymbol{x}) = \sum_{i=1}^{n} \mathcal{I}(x_i \neq 0)$$

其中，

$$\mathcal{I}(x_i \neq 0) = \begin{cases} 1 & , x_i \neq 0 \\ 0 & , x_i = 0 \end{cases}$$

基数函数也被称为 l_0 范数，但是它并不满足范数定义的条件。

例 7

求向量 $\boldsymbol{x}=(-1,2,4)^{\mathrm{T}}$ 的 0，1，2，和 ∞-范数。

解

$$
\begin{aligned}
||\boldsymbol{x}||_0 &= 3\\
||\boldsymbol{x}||_1 &= |-1|+2+4=7\\
||\boldsymbol{x}||_2 &= \sqrt{|-1|^2+2^2+4^2}=\sqrt{21}\\
||\boldsymbol{x}||_\infty &= \max\{|-1|,2,4\}=4
\end{aligned}
$$

6.1.3 范数的几何意义：单位范数球

定义 3

对于 l_p 范数小于等于 *1* 的向量集合，

$$\mathcal{B}_p = \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\|_p \leq 1\}$$

称为 l_p 的**单位范数球**。

范数的几何意义

例 8

单位范数球的形状反映了不同范数的性质，对于不同的 p，范数球有着不同的几何形状。图2分别表示了 $\mathcal{B}_2, \mathcal{B}_1, \mathcal{B}_\infty$ 在 $\mathbb{R}^2$ 的范数球形状。

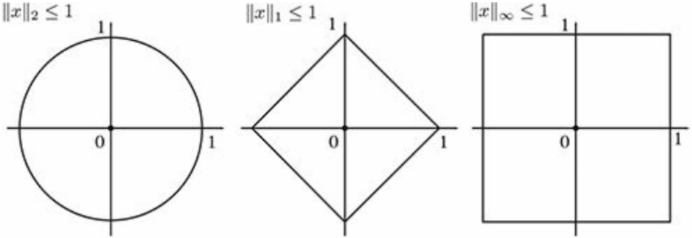

图 2: $\mathbb{R}^2$ 上的范数球

6.1.4 范数性质

定义 4

设 $\{\boldsymbol{x}^{(k)}\}$ 为 $\mathbb{R}^n$ 中一向量序列，$\boldsymbol{x}^* \in \mathbb{R}^n$，其中

$$\boldsymbol{x}^{(k)} = (x_1^{(k)}, x_2^{(k)}, \cdots, x_n^{(k)})^{\mathrm{T}}, \boldsymbol{x}^* = (x_1^*, x_2^*, \cdots, x_n^*)^{\mathrm{T}}$$

如果 $\lim_{k\to\infty} x_i^{(k)} = x_i^*(i = 1, 2, \cdots, n)$，则称 $\boldsymbol{x}^{(k)}$ 收敛于向量 x^*，记作

$$\lim_{k\to\infty} \boldsymbol{x}^{(k)} = \boldsymbol{x}^*$$

范数的连续性和收敛性质

定理 1

（范数的连续性）设非负函数 $N(\boldsymbol{x}) = \|\boldsymbol{x}\|$ 为 $\mathbb{R}^n$ 上任一向量范数，则 $N(\boldsymbol{x})$ 是 $\boldsymbol{x}$ 分量 $x_1, x_2, \cdots, x_n$ 的连续函数。

定理 2

（柯西收敛原理）实数域 $\mathbb{R}$（或者复数域 $\mathbb{C}$）上的有限维线性空间按任何范数 $\|\cdot\|$ 必定完备。

定理 3

（向量序列收敛定理）设 $\{\boldsymbol{x}^{(k)}\}$ 为 $\mathbb{R}^n$ 中一向量序列，$\boldsymbol{x}^* \in \mathbb{R}^n$ 则

$$\lim_{k\to\infty} \boldsymbol{x}^{(k)} = \boldsymbol{x}^* \Longleftrightarrow \lim_{k\to\infty} \left\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\right\| = 0$$

其中 $\|\cdot\|$ 为向量的任一种范数。

范数的等价性

在 $\mathbb{R}^n$（或 $\mathbb{C}^n$）上可以定义各种向量范数，其数值大小一般不同，但是在各种向量范数之间存在下述重要的关系

$$\|\boldsymbol{x}\|_\infty \le \|\boldsymbol{x}\|_1 \le n\|\boldsymbol{x}\|_\infty$$

$$\frac{1}{\sqrt{n}}\|\boldsymbol{x}\|_1 \le \|\boldsymbol{x}\|_2 \le \|\boldsymbol{x}\|_1$$

$$\frac{1}{\sqrt{n}}\|\boldsymbol{x}\|_2 \le \|\boldsymbol{x}\|_\infty \le \|\boldsymbol{x}\|_2$$

或者

$$\|\boldsymbol{x}\|_\infty \le \|\boldsymbol{x}\|_2 \le \|\boldsymbol{x}\|_1 \le \sqrt{n}\|\boldsymbol{x}\|_2 \le n\|\boldsymbol{x}\|_\infty$$

范数的等价性

定理 4

（向量范数的等价性定理）设 $\|\boldsymbol{x}\|_s, \|\boldsymbol{x}\|_t$ 为 $\mathbb{R}^n$ 上的任意两种向量范数，则存在两个与向量无关的正常数 $c_1, c_2 > 0$，使得下面的不等式成立

$$c_1 \|\boldsymbol{x}\|_s \leq \|\boldsymbol{x}\|_t \leq c_2 \|\boldsymbol{x}\|_s, \quad \text{对一切 } \boldsymbol{x} \in \mathbb{R}^n$$

并称 $\|\boldsymbol{x}\|_t$ 和 $\|\boldsymbol{x}\|_s$ 为 $\mathbb{R}^n$ 上的等价范数。