zhuanlan.zhihu.com

在众多的机器学习模型中,线性代数的身影无处不在,当然,我们也会时常碰到线性代数中的正定矩阵 和半正定矩阵。例如,多元正态分布的协方差矩阵要求是半正定的。

1. 基本的定义

正定和半正定这两个词的英文分别是 positive definite 和 positive semi-definite, 其中, definite 是一个形容词,表示"明确的、确定的"等意思。

初学线性代数的读者可能会被这两个词 "唬住",但正定矩阵和半正定矩阵的定义实际上是很简单的 (不考虑复数构成的矩阵):

【定义 1】给定一个大小为 $n \times n$ 的实对称矩阵 A ,若对于任意长度为 $m{n}$ 的非零向量 $m{x}$,有 $m{x}^T A m{x} > m{0}$ 恒成立,则矩阵 A 是一个正定矩阵。

【例 1】单位矩阵 $I \in \mathbb{R}^{2 \times 2}$ 是否是正定矩阵?

解: 设向量
$$oldsymbol{x} = \left[egin{array}{c} x_1 \ x_2 \end{array}
ight] \in \mathbb{R}^2$$
 为非零向量,则

$$oldsymbol{x}^T I oldsymbol{x} = oldsymbol{x}^T oldsymbol{x} = oldsymbol{x}_1^2 + oldsymbol{x}_2^2$$

由于 $oldsymbol{x}
eq oldsymbol{0}$,故 $oldsymbol{x}^T I oldsymbol{x} > oldsymbol{0}$ 恒成立,即单位矩阵 $I \in \mathbb{R}^{2 imes 2}$ 是正定矩阵。

单位矩阵是正定矩阵 (positive definite)。

【简单证明】对于任意单位矩阵 $I \in \mathbb{R}^{n imes n}$ 而言,给定任意非零向量 $oldsymbol{x} \in \mathbb{R}^n$,恒有

$$oldsymbol{x}^T I oldsymbol{x} = oldsymbol{x}^T oldsymbol{x} \ = x_1^2 + x_2^2 + \dots + x_n^2 > 0$$

【例 2】实对称矩阵
$$A=egin{bmatrix}2&-1&0\-1&2&-1\0&-1&2\end{bmatrix}\in\mathbb{R}^{3 imes3}$$
 是否是正定矩阵?

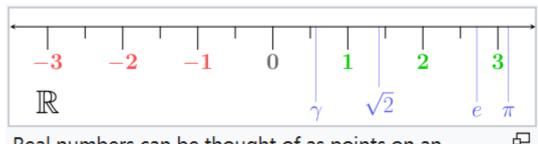
解: 设向量
$$oldsymbol{x} = egin{bmatrix} x_1 \ x_2 \ x_3 \end{bmatrix} \in \mathbb{R}^3$$
 为非零向量,则

$$egin{align} m{x}^T A m{x} &= \left[\, (2x_1 - x_2) \quad (-x_1 + 2x_2 - x_3) \quad -x_2 + 2x_3 \,
ight] egin{bmatrix} x_1 \ x_2 \ x_3 \ \end{bmatrix} \ \ &= x_1^2 + (x_1 - x_2)^2 + (x_2 - x_3)^2 + x_3^2 > 0 \ \end{aligned}$$

因此, 矩阵 A 是正定矩阵。

【定义 2】给定一个大小为 $n \times n$ 的实对称矩阵 A ,若对于任意长度为 $m{n}$ 的向量 $m{x}$,有 $m{x}^T A m{x} \geq \mathbf{0}$ 恒成立,则矩阵 A 是一个半正定矩阵。

根据正定矩阵和半正定矩阵的定义,我们也会发现:半正定矩阵包括了正定矩阵,与非负实数 (nonnegative real number) 和正实数 (positive real number) 之间的关系很像。

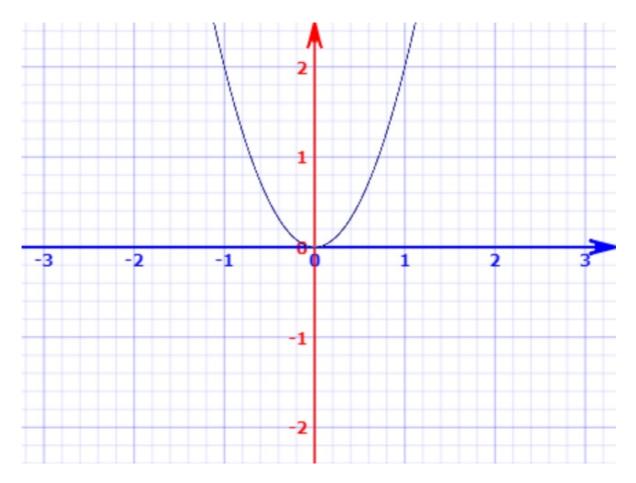


Real numbers can be thought of as points on an infinitely long number line

2. 从二次函数到正定 / 半正定矩阵

在初中数学中,我们学习了二次函数 $y=ax^2$,该函数的曲线会经过坐标原点,当参数 a>0时,曲线的 "开口" 向上,参数 a<0时,曲线的 "开口" 向下。

以 $y=2x^2$ 为例,曲线如下:



实际上,我们可以将 $oldsymbol{y} = oldsymbol{x}^T A oldsymbol{x}$ 视作 $oldsymbol{y} = a oldsymbol{x}^2$ 的多维表达式。

当我们希望 $m{y}=m{x}^T A m{x} \geq 0$ 对于任意向量 $m{x}$ 都恒成立,就要求矩阵 A 是一个半正定矩阵,对应于二次函数, $m{y}=a m{x}^2 > 0, orall m{x}$ 需要使得 $a \geq 0$.

另外,在 $y=ax^2$ 中,我们还知道:若 a>0 ,则对于任意 x
eq 0 ,有 y>0 恒成立。

这在 $oldsymbol{y}=oldsymbol{x}^T A oldsymbol{x}$ 也有契合之处,当矩阵 A 是正定矩阵时,对于任意 $oldsymbol{x}
eq oldsymbol{0}$, $oldsymbol{y}>oldsymbol{0}$ 恒成 立。

3. 正定矩阵和半正定矩阵的直观解释

若给定任意一个正定矩阵 $A\in\mathbb{R}^{n imes n}$ 和一个非零向量 $m{x}\in\mathbb{R}^n$,则两者相乘得到的向量 $m{y}=Am{x}\in\mathbb{R}^n$ 与向量 $m{x}$ 的夹角恒小于 $\dfrac{\pi}{2}$. (等价于: $m{x}^TAm{x}>0$.)

【例 3】给定向量
$$m{x}=egin{bmatrix}2\\1\end{bmatrix}\in\mathbb{R}^2$$
 ,对于单位矩阵 $m{I}=egin{bmatrix}1&0\\0&1\end{bmatrix}\in\mathbb{R}^{2 imes2}$,则

$$oldsymbol{y} = Ioldsymbol{x} = oldsymbol{x} = egin{bmatrix} 2 \ 1 \end{bmatrix}$$

向量 $oldsymbol{x},oldsymbol{y}\in\mathbb{R}^2$ 之间的夹角为

$$egin{align} \cos \langle oldsymbol{x}, oldsymbol{y}
angle &= rac{oldsymbol{x}^T oldsymbol{y}}{||oldsymbol{x}|| \cdot ||oldsymbol{y}||} \ &= rac{2 imes 2 + 1 imes 1}{\sqrt{2^2 + 1^2} \cdot \sqrt{2^2 + 1^2}} \ &= 1 \end{split}$$

即两个向量之间的夹角为 0°.

【例 4】给定向量
$$oldsymbol{x}=egin{bmatrix}1\\2\\1\end{bmatrix}\in\mathbb{R}^3$$
 ,对于实对称矩阵 $A=egin{bmatrix}2&-1&0\\-1&2&-1\\0&-1&2\end{bmatrix}\in\mathbb{R}^{3 imes3}$,则

$$oldsymbol{y} = Aoldsymbol{x} = egin{bmatrix} 0 \ 2 \ 0 \end{bmatrix}$$

向量 $oldsymbol{x},oldsymbol{y}\in\mathbb{R}^2$ 之间的夹角为

$$\cos\langle oldsymbol{x}, oldsymbol{y}
angle = rac{oldsymbol{x}^Toldsymbol{y}}{||oldsymbol{x}||\cdot||oldsymbol{y}||} = rac{\sqrt{6}}{3}$$

即两个向量之间的夹角小于 $\frac{\pi}{2}$.

若给定任意一个半正定矩阵 $A\in\mathbb{R}^{n imes n}$ 和一个向量 $m{x}\in\mathbb{R}^n$,则两者相乘得到的向量 $m{y}=Am{x}\in\mathbb{R}^n$ 与向量 $m{x}$ 的夹角恒小于或等于 $\dfrac{\pi}{2}$. (等价于: $m{x}^TAm{x}\geq 0$.)

3. 为什么协方差矩阵要是半正定的?

在概率论与数理统计中, 我们都学习的协方差矩阵的定义:

对于任意多元随机变量
$$oldsymbol{t}$$
 ,协方差矩阵为 $C=\mathbb{E}\left[(oldsymbol{t}-ar{oldsymbol{t}})(oldsymbol{t}-ar{oldsymbol{t}})^T
ight]$

现给定任意一个向量 农,则

$$egin{align} oldsymbol{x}^T C oldsymbol{x} &= oldsymbol{x}^T \mathbb{E}\left[(oldsymbol{t} - ar{oldsymbol{t}})(oldsymbol{t} - ar{oldsymbol{t}})^T oldsymbol{x}
ight] \ &= \mathbb{E}(oldsymbol{s}^2) = \sigma_s^2 \end{aligned}$$

其中,

$$\sigma_s = oldsymbol{x}^T (oldsymbol{t} - ar{oldsymbol{t}}) = (oldsymbol{t} - ar{oldsymbol{t}})^T oldsymbol{x}$$

由于 $\sigma_s^2 \geq 0$,因此, $oldsymbol{x}^T C oldsymbol{x} \geq 0$, 协方差矩阵 C 是半正定的。

相关参考:

Positive-definite matrixen.wikipedia.org