统计方法与机器学习

第〇章：基础回顾

倪 葎

DaSE@ECNU
(lni@dase.ecnu.edu.cn)

目录

向量

在实数域 R^n 上的一个 n 维向量

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = (a_1, a_2, \cdots, a_n)'.$$

常用运算

- **加法** 设另有一个 n 维向量 $\boldsymbol{b} = (b_1, b_2, \cdots, b_n)'$，则

$$\boldsymbol{a} + \boldsymbol{b} = (a_1 + b_1, a_2 + b_2, \cdots, a_n + b_n)'$$

- **数乘**设 c 为一个实数，则

$$c\boldsymbol{a} = (ca_1, ca_2, \cdots, ca_n)'$$

向量

常用运算性质

- 若 $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in R^n$，则

$$\boldsymbol{a}+\boldsymbol{b}=\boldsymbol{b}+\boldsymbol{a}, \quad (\boldsymbol{a}+\boldsymbol{b})+\boldsymbol{c}=\boldsymbol{a}+(\boldsymbol{b}+\boldsymbol{c})$$
$$\boldsymbol{a}+\boldsymbol{0}=\boldsymbol{a}, \quad \boldsymbol{a}+(-\boldsymbol{a})=\boldsymbol{0},$$

其中 $\boldsymbol{0}=(0,0,\cdots,0)'$.

- 若 $\boldsymbol{a}, \boldsymbol{b} \in R^n, c, c_1, c_2 \in R$，则

$$c(\boldsymbol{a}+\boldsymbol{b})=c\boldsymbol{a}+c\boldsymbol{b}, \quad (c_1+c_2)\boldsymbol{a}=c_1\boldsymbol{a}+c_2\boldsymbol{a}$$
$$c_1(c_2\boldsymbol{a})=c_1c_2\boldsymbol{a}, \quad 1\cdot\boldsymbol{a}=\boldsymbol{a}.$$

线性相关

- 对于向量 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k$，若存在一组不全为零常数 $c_1, c_2, \cdots, c_k$ 使得

$$c_1\boldsymbol{a}_1 + c_2\boldsymbol{a}_2 + \cdots + c_k\boldsymbol{a}_k = \mathbf{0},$$

  则称向量 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k$ 是线性相关的；否则，称 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k$ 为线性无关。

向量

任意两个向量 $\boldsymbol{a} = (a_1, a_2, \cdots, a_p)'$ 和 $\boldsymbol{b} = (b_1, b_2, \cdots, b_p)'$。$\boldsymbol{a}, \boldsymbol{b}$ 的内积为

$$\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \boldsymbol{a}'\boldsymbol{b} = \sum_{i=1}^{p} a_i b_i.$$

内积的性质

- $\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \boldsymbol{a}'\boldsymbol{b} = \boldsymbol{b}'\boldsymbol{a} = \langle \boldsymbol{b}, \boldsymbol{a} \rangle$；
- $\langle \boldsymbol{a}, \boldsymbol{a} \rangle = \boldsymbol{a}'\boldsymbol{a} \geq 0$，当且仅当 $\boldsymbol{a} = \boldsymbol{0}$；记 $\langle \boldsymbol{a}, \boldsymbol{a} \rangle = \|\boldsymbol{a}\|^2$；
- $\langle c\boldsymbol{a}, \boldsymbol{b} \rangle = \langle \boldsymbol{a}, c\boldsymbol{b} \rangle = c\langle \boldsymbol{a}, \boldsymbol{b} \rangle$ 对一切 $c \in R$ 成立；
- 对于三个向量 $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$，有

$$\langle \boldsymbol{a}, \boldsymbol{b} + \boldsymbol{c} \rangle = \langle \boldsymbol{a}, \boldsymbol{b} \rangle + \langle \boldsymbol{a}, \boldsymbol{c} \rangle.$$

内积的不等式

- 柯西不等式：$\langle \boldsymbol{a}, \boldsymbol{b} \rangle^2 \leq \|\boldsymbol{a}\|^2 \cdot \|\boldsymbol{b}\|^2$;
- 三角不等式：$\|\boldsymbol{a} + \boldsymbol{b}\| \leq \|\boldsymbol{a}\| + \|\boldsymbol{b}\|$.

生成子空间

- 给定向量 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k$，考虑由这些向量所有可能的线性组合 $\sum_{i=1}^{k} c_i \boldsymbol{a}_i$ 组成的集合

$$\mathcal{L}(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k) = \left\{ \sum_{i=1}^{k} c_i \boldsymbol{a}_i : c_1, c_2, \cdots, c_k \in R \right\}.$$

称其是由向量 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k$ **生成子空间**。

基

- 设 $\mathcal{L}$ 是 R^n 中的一个子空间，如果存在 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k$ 使得
$$\mathcal{L} = \mathcal{L}(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k)$$
且 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k$ 线性无关，则称 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k$ 是 $\mathcal{L}$ 的一组**基**。
- 根据定义，如果 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k$ 是子空间 $\mathcal{L}$ 的一组基，那么 $\mathcal{L}$ 中任一向量 $\boldsymbol{a}$ 都可被 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k$ 的线性组合来表示，而且这种表示法是**唯一**的。
- 可以证明，子空间 $\mathcal{L}$ 中如有两组基，那么这两组基中向量的个数一定相同。
- 子空间 $\mathcal{L}$ 中一组基所含的向量的个数成为 $\mathcal{L}$ 的**维数**。

标准正交基

- 在一个 R^n 的子空间 $\mathcal{L}$ 的基 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k$ 具有性质

$$\begin{aligned} \langle \boldsymbol{a}_i, \boldsymbol{a}_i \rangle &= 1, i = 1, 2, \cdots, k, \\ \langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle &= 0, i \neq j. \end{aligned}$$

则称 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k$ 是 $\mathcal{L}$ 的一组标准正交基。

向量

在 R^n 中，给定一个向量 $\boldsymbol{a}$ 及子空间 $\mathcal{L}$，如果在 $\mathcal{L}$ 中存在 $\boldsymbol{b}$ 使

$$\|\boldsymbol{a}-\boldsymbol{b}\| = \inf_{\boldsymbol{x}\in\mathcal{L}} \|\boldsymbol{a}-\boldsymbol{x}\|$$

则称 $\boldsymbol{b}$ 是 $\boldsymbol{a}$ 在 $\mathcal{L}$ 中的投影。

投影的性质

- 投影是存在且唯一的。

引理

在 R^n 中，给定一个向量 $\boldsymbol{a}$ 及子空间 $\mathcal{L}$，$\boldsymbol{b}$ 是 $\boldsymbol{a}$ 在 $\mathcal{L}$ 中的投影，当且仅当

$$\langle \boldsymbol{a}-\boldsymbol{b}, \boldsymbol{x}\rangle = 0, \quad \text{对一切 } \boldsymbol{x}\in\mathcal{L} \text{ 成立。}$$

证明

$(\Rightarrow)$ 采用反证法证明。
设对于任意 $\boldsymbol{x} \in \mathcal{L}$ 使得 $\langle \boldsymbol{a}-\boldsymbol{b}, \boldsymbol{x}\rangle \neq 0$。对于一切的 λ，有

$$
\begin{aligned}
& \langle \boldsymbol{a}-\boldsymbol{b}, \boldsymbol{a}-\boldsymbol{b}\rangle \\
\leq\ & \langle \boldsymbol{a}-\boldsymbol{b}+\lambda \boldsymbol{x}, \boldsymbol{a}-\boldsymbol{b}+\lambda \boldsymbol{x}\rangle \quad \text{（投影的定义）} \\
=\ & \langle \boldsymbol{a}-\boldsymbol{b}, \boldsymbol{a}-\boldsymbol{b}\rangle + 2\lambda\langle \boldsymbol{a}-\boldsymbol{b}, \boldsymbol{x}\rangle + \lambda^2\langle \boldsymbol{x}, \boldsymbol{x}\rangle
\end{aligned}
$$

只要选 $\lambda = -\varepsilon\langle \boldsymbol{a}-\boldsymbol{b}, \boldsymbol{x}\rangle, \varepsilon > 0$，上式就可写成

$$\langle \boldsymbol{a}-\boldsymbol{b}, \boldsymbol{x}\rangle^2(\varepsilon^2\|\boldsymbol{x}\|^2 - 2\varepsilon) \geq 0$$

对一切的 $\varepsilon > 0$ 都成立，这是不可能的。因此，

$$\langle \boldsymbol{a}-\boldsymbol{b}, \boldsymbol{x}\rangle = 0.$$

证明

$(\Leftarrow)$ 由于

$$\begin{aligned} & \langle \boldsymbol{a}-\boldsymbol{x}, \boldsymbol{a}-\boldsymbol{x}\rangle \\ = & \langle \boldsymbol{a}-\boldsymbol{b}+\boldsymbol{b}-\boldsymbol{x}, \boldsymbol{a}-\boldsymbol{b}+\boldsymbol{b}-\boldsymbol{x}\rangle \\ = & \langle \boldsymbol{a}-\boldsymbol{b}, \boldsymbol{a}-\boldsymbol{b}\rangle+2\langle \boldsymbol{a}-\boldsymbol{b}, \boldsymbol{b}-\boldsymbol{x}\rangle+\langle \boldsymbol{b}-\boldsymbol{x}, \boldsymbol{b}-\boldsymbol{x}\rangle, \end{aligned}$$

注意到，$\boldsymbol{b} \in \mathcal{L}, \boldsymbol{x} \in \mathcal{L}$，则 $\boldsymbol{b}-\boldsymbol{x} \in \mathcal{L}$。
于是，$\langle \boldsymbol{a}-\boldsymbol{b}, \boldsymbol{b}-\boldsymbol{x}\rangle = 0$。所以，由

$$\langle \boldsymbol{a}-\boldsymbol{x}, \boldsymbol{a}-\boldsymbol{x}\rangle = \langle \boldsymbol{a}-\boldsymbol{b}, \boldsymbol{a}-\boldsymbol{b}\rangle+\langle \boldsymbol{b}-\boldsymbol{x}, \boldsymbol{b}-\boldsymbol{x}\rangle$$

可知，

$$\boldsymbol{b} = \arg\min_{x\in\mathcal{L}} \|\boldsymbol{a}-\boldsymbol{x}\|.$$

因此，$\boldsymbol{b}$ 是 $\boldsymbol{a}$ 在 $\mathcal{L}$ 中的投影。

格兰姆-施密特正交化方法

- 目的：任意一组线性无关的向量 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k$，可以找到子空间 $\mathcal{L}(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k)$ 中的一组标准正交基。

- 具体步骤：
  - 令 $\boldsymbol{b}_1 = \boldsymbol{a}_1$;
  - 令 $\boldsymbol{b}_i = \boldsymbol{a}_i - h_{ii-1}\boldsymbol{b}_{i-1} - \cdots - h_{i1}\boldsymbol{b}_1$ 使得

    $$\langle \boldsymbol{b}_i, \boldsymbol{b}_{i-1} \rangle = \langle \boldsymbol{b}_i, \boldsymbol{b}_{i-2} \rangle = \cdots = \langle \boldsymbol{b}_i, \boldsymbol{b}_1 \rangle = 0,$$

    确定系数 $h_{ii-1}, \cdots, h_{i1}$
  - 以此类推，确定一组两两正交的向量 $\boldsymbol{b}_1, \cdots, \boldsymbol{b}_k$;
  - 令 $\boldsymbol{e}_i = \|\boldsymbol{b}_i\|^{-1}\boldsymbol{b}_i, i = 1, 2, \cdots, k$, 则 $\boldsymbol{e}_1, \cdots, \boldsymbol{b}_k$ 就是标准正交的向量组。

直接和与正交和

设 $\mathcal{L}$ 是一个子空间，$\mathcal{L}_1, \cdots, \mathcal{L}_k$ 是 k 个子空间。

- 对每一 $\boldsymbol{a} \in \mathcal{L}$, 如果将 $\boldsymbol{a}$ 能唯一地表示为 $\boldsymbol{a}_1+\boldsymbol{a}_2+\cdots+\boldsymbol{a}_k$，其中 $\boldsymbol{a}_i \in \mathcal{L}_i, i=1,2,\cdots,k$，则称 $\mathcal{L}$ 是 $\mathcal{L}_1, \cdots, \mathcal{L}_k$ 的**直接和**，记 $\mathcal{L}=\mathcal{L}_1+\cdots+\mathcal{L}_k$.
- 若 $\mathcal{L}$ 是 $\mathcal{L}_1, \cdots, \mathcal{L}_k$ 的直接和，并且当 $i \neq j$ 时只要 $\boldsymbol{a}_i \in \mathcal{L}_i, \boldsymbol{a}_j \in \mathcal{L}_j$ 有 $\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle = 0$,那么称 $\mathcal{L}$ 是 $\mathcal{L}_1, \cdots, \mathcal{L}_k$ 的**正交和**，记
$$\mathcal{L}=\mathcal{L}_1 \oplus \cdots \oplus \mathcal{L}_k.$$
- 特别地，如果 $R_n=\mathcal{L}_1 \oplus \mathcal{L}_2$，则称 $\mathcal{L}_1(\mathcal{L}_2)$ 是 $\mathcal{L}_2(\mathcal{L}_1)$ 的正交补空间。

矩阵

称

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

为大小 $m \times n$ 的矩阵 $\boldsymbol{A} = \{a_{ij}\}_{m\times n}$。

常用符号

- $\boldsymbol{I}$：单位矩阵，主对角线的元素为一，其他元素为零；
- $\boldsymbol{A}' = \{a_{ji}\}_{n\times m}$：转置；
- $\mathrm{rank}(\boldsymbol{X})$：秩；
    - $\mathrm{rank}(\boldsymbol{A}) = \mathrm{rank}(\boldsymbol{A}')$；
    - $\mathrm{rank}(\boldsymbol{AB}) \le \min\{\mathrm{rank}(\boldsymbol{A}), \mathrm{rank}(\boldsymbol{B})\}$；

矩阵

当矩阵 $\boldsymbol{A}$ 的行数等于列数时，也称 $\boldsymbol{A}$ 为方阵。

常用符号

- $\boldsymbol{A}^{-1}$：逆矩阵；
- $\boldsymbol{A}^{*}$：伴随矩阵；
- $|\boldsymbol{A}|$：行列式；
  - $|\boldsymbol{A}| = |\boldsymbol{A}'|$；
  - 当 $|\boldsymbol{A}| \neq 0$ 时，$|\boldsymbol{A}|\boldsymbol{A}^{-1} = \boldsymbol{A}^{*}$；
  - $|\boldsymbol{AB}| = |\boldsymbol{A}||\boldsymbol{B}|$；
  - $|\boldsymbol{A}| \neq 0 \Leftrightarrow \boldsymbol{A}_{n\times n}$ 非奇异 $\Leftrightarrow \mathrm{rank}(\boldsymbol{A}) = n$；
  - $|\boldsymbol{A}|^{-1} = |\boldsymbol{A}^{-1}|$。

分块矩阵的逆

$$\boldsymbol{A}=\begin{pmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{pmatrix}$$

- 若 $\boldsymbol{A}^{-1}$ 存在时，且 $|\boldsymbol{A}_{11}|\neq 0$，则

$$\begin{aligned}&\begin{pmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{pmatrix}^{-1}\\ =&\begin{pmatrix}\boldsymbol{A}_{11}^{-1}+\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{M}^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{M}^{-1}\\ -\boldsymbol{M}^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \boldsymbol{M}^{-1}\end{pmatrix}\end{aligned}$$

其中 $\boldsymbol{M}=\boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}$。

分块矩阵的行列式

$$\boldsymbol{A}=\begin{pmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{pmatrix}$$

- 若 $|\boldsymbol{A}_{11}|\neq 0$，则

$$\begin{vmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{vmatrix}=|\boldsymbol{A}_{11}||\boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}|.$$

- 若 $|\boldsymbol{A}_{22}|\neq 0$，则

$$\begin{vmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{vmatrix}=|\boldsymbol{A}_{22}||\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}|.$$

特征根和特征向量

给定一个 $n \times n$ 的方阵 $\boldsymbol{A}$,

- $\lambda \boldsymbol{I} - \boldsymbol{A}$ 是 λ 的 n 次多项式，称为 $\boldsymbol{A}$ 的特征多项式。
- $|\lambda \boldsymbol{I} - \boldsymbol{A}| = 0$ 称为 $\boldsymbol{A}$ 的特征方程。
- 特征方程的解（或特征多项式的根），称为 $\boldsymbol{A}$ 的特征根。
- 因为 $|\lambda \boldsymbol{I} - \boldsymbol{A}| = 0$，所以方程

$$(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$$

一定有非零解，λ 所对应的非 0 解向量称为 λ 相应的特征向量，也称为 $\boldsymbol{A}$ 的特征向量。

迹

对于 $n \times n$ 的方阵 $\boldsymbol{A}$，$\boldsymbol{A}$ 的迹 $\mathrm{tr}(\boldsymbol{A})$ 是 $\boldsymbol{A}$ 的所有特征根之和，即若 $\boldsymbol{A}$ 的特征根为 $\lambda_1, \lambda_2, \cdots, \lambda_n$，则

$$\mathrm{tr}(\boldsymbol{A}) = \sum_{i=1}^{n} \lambda_i.$$

- $\mathrm{tr}(\boldsymbol{A}+\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}) + \mathrm{tr}(\boldsymbol{B})$；
- $\mathrm{tr}(c\boldsymbol{A}) = c\,\mathrm{tr}(\boldsymbol{A})$；
- 当 $\boldsymbol{AB}$ 和 $\boldsymbol{BA}$ 均为方阵（但大小不要求相同）时，有

$$\mathrm{tr}(\boldsymbol{AB}) = \mathrm{tr}(\boldsymbol{BA}).$$

  - 特例：$\mathrm{tr}(\boldsymbol{A}'\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{A}\boldsymbol{A}')$。
  - 特例：$\mathrm{tr}(\boldsymbol{a}\boldsymbol{b}') = \mathrm{tr}(\boldsymbol{b}'\boldsymbol{a}) = \boldsymbol{b}'\boldsymbol{a}$。

特征根与行列式的关系

若 $\boldsymbol{A}$ 的特征根为 $\lambda_1, \lambda_2, \cdots, \lambda_n$，则 $\boldsymbol{A}$ 的行列式为

$$|\boldsymbol{A}| = \prod_{i=1}^{n} \lambda_i.$$

- $\boldsymbol{A}$ 非奇异，当且仅当 $\boldsymbol{A}$ 的特征根均不为零；
- $\boldsymbol{A}$ 奇异，当且仅当 $\boldsymbol{A}$ 至少有一个特征根为零；

幂等阵

如果方阵 $\boldsymbol{A}$ 具有性质 $\boldsymbol{A}^2 = \boldsymbol{A}$，则称 $\boldsymbol{A}$ 是幂等阵。

- 如果 $\boldsymbol{A}^2 = \boldsymbol{A}$，则 $\boldsymbol{A}$ 的特征根非零即一。
- 如果 $\boldsymbol{A}^2 = \boldsymbol{A}$，则 $\mathrm{rank}(\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{A})$。
- A 是幂等阵，当且仅当

$$R^n = \mathcal{L}(\boldsymbol{A}') \oplus \mathcal{L}(\boldsymbol{I} - \boldsymbol{A}) = \mathcal{L}(\boldsymbol{A}) \oplus \mathcal{L}(\boldsymbol{I} - \boldsymbol{A}').$$

对称阵

如果方阵 $\boldsymbol{A}$ 具有性质 $\boldsymbol{A}'=\boldsymbol{A}$，则称 $\boldsymbol{A}$ 是对称阵。

- 对每个对称阵 $\boldsymbol{A}$，任给一个向量 $\boldsymbol{x}$，$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ 是 $\boldsymbol{A}$ 的一个齐次二次函数，称为 $\boldsymbol{A}$ 对应的二次型。
- 如果 $\boldsymbol{A}$ 的二次型 $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ 恒不取负值，即 $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}\geq 0$ 对一切 $\boldsymbol{x}$ 成立，则称 $\boldsymbol{A}$ 是非负定阵。
- 如果 $\boldsymbol{A}$ 是非负定阵，且 $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}=0$ 充要条件是 $\boldsymbol{x}=0$，则称 $\boldsymbol{A}$ 是正定的。

对称阵的谱分解

- 对于对称阵 $\boldsymbol{A}$，设特征根分别为 $\lambda_1, \lambda_2, \cdots, \lambda_n$，可证明 λ_i 是实数，$i = 1, 2, \cdots, n$。记

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

- 设 $\boldsymbol{A}$ 的 n 个单位正交特征向量为 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$，记

$$\boldsymbol{V}' = (\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n)$$

可证明 $\boldsymbol{V}'\boldsymbol{V} = \boldsymbol{V}\boldsymbol{V}' = \boldsymbol{I}_n$。于是，$\boldsymbol{A}\boldsymbol{V}' = \boldsymbol{V}'\boldsymbol{\Lambda}$。

对称阵的性质

设 $\boldsymbol{A}$ 是一个对称阵，$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ 是 $\boldsymbol{A}$ 的特征根，而 $\boldsymbol{v}_i$ 是 λ_i 对应的特征向量。于是，有以下结论：

- 对于任意非零向量 $\boldsymbol{x}$，

$$\lambda_n \boldsymbol{x}'\boldsymbol{x} \leq \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \leq \lambda_1 \boldsymbol{x}'\boldsymbol{x}.$$

- 推论：对于任意非零向量 $\boldsymbol{x}$，

$$\sup_{\boldsymbol{x}} \frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}} = \lambda_1, \quad \inf_{\boldsymbol{x}} \frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}} = \lambda_n,$$

证明：仅考虑 $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \le \lambda_1 \boldsymbol{x}'\boldsymbol{x}$

不妨设特征向量 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$ 是对称阵 $\boldsymbol{A}$ 的标准正交基。因此，

$$\boldsymbol{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i, \quad \boldsymbol{v}_j'\boldsymbol{v}_i = \begin{cases} 0, i \ne j \\ 1, i = j \end{cases}.$$

对于任意非零向量 $\boldsymbol{x}$，$\boldsymbol{x}$ 可以写成 $\boldsymbol{x} = \sum_{i=1}^n k_i \boldsymbol{v}_i$。

证明：仅考虑 $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \leq \lambda_1 \boldsymbol{x}'\boldsymbol{x}$ (续)

于是有

$$
\begin{aligned}
\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} &= \left(\sum_{j=1}^{n} k_j \boldsymbol{v}_j\right)' \boldsymbol{A} \left(\sum_{i=1}^{n} k_i \boldsymbol{v}_i\right) \\
&= \left(\sum_{j=1}^{n} k_j \boldsymbol{v}_j\right)' \left(\sum_{i=1}^{n} \lambda_i k_i \boldsymbol{v}_i\right) \\
&= \sum_{i,j} \lambda_i k_i k_j \boldsymbol{v}_j' \boldsymbol{v}_i \\
&= \sum_{i} \lambda_i k_i^2 \\
&\leq \lambda_1 \sum_{i} k_i^2 = \lambda_1 \boldsymbol{x}'\boldsymbol{x}.
\end{aligned}
$$

求导法则

- **列向量对标量求导**：若

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},$$

x 是一个标量，则其微商也是一个向量，即

$$\frac{\partial \boldsymbol{y}}{\partial x} = \begin{pmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \\ \vdots \\ \frac{\partial y_n}{\partial x} \end{pmatrix}$$

求导法则

- **矩阵对标量求导**：若

$$
\boldsymbol{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{pmatrix},
$$

x 是一个标量，则其微商也是一个 $m \times n$ 矩阵，即

$$
\frac{\partial \boldsymbol{Y}}{\partial x} = \begin{pmatrix} \frac{\partial y_{11}}{\partial x} & \frac{\partial y_{12}}{\partial x} & \cdots & \frac{\partial y_{1n}}{\partial x} \\ \frac{\partial y_{21}}{\partial x} & \frac{\partial y_{22}}{\partial x} & \cdots & \frac{\partial y_{2n}}{\partial x} \\ \vdots & \vdots & & \vdots \\ \frac{\partial y_{m1}}{\partial x} & \frac{\partial y_{m2}}{\partial x} & \cdots & \frac{\partial y_{mn}}{\partial x} \end{pmatrix}
$$

求导法则

- **标量对向量求导**：若

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

y 是一个标量，则其微商也是一个向量，即

$$\frac{\partial y}{\partial \boldsymbol{x}} = \begin{pmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{pmatrix}$$

求导法则

- **标量对矩阵求导**：若

$$\boldsymbol{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix},$$

y 是一个标量，则其微商也是一个矩阵，即

$$\frac{\partial y}{\partial \boldsymbol{X}} = \begin{pmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{12}} & \cdots & \frac{\partial y}{\partial x_{1n}} \\ \frac{\partial y}{\partial x_{21}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{2n}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial y}{\partial x_{m1}} & \frac{\partial y}{\partial x_{m2}} & \cdots & \frac{\partial y}{\partial x_{mn}} \end{pmatrix}.$$

求导法则

- **列向量对行向量求导**：若 $\boldsymbol{y} = (y_1, y_2, \cdots, y_m)'$ 是一个 $m \times 1$ 列向量，而 $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ 是一个 $n \times 1$ 列向量，则

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}'} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{pmatrix}$$

是一个 $m \times n$ 矩阵。

求导法则

- **行向量对列向量求导**：若 $\boldsymbol{y} = (y_1, y_2, \cdots, y_m)'$ 是一个 $m \times 1$ 列向量，而 $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ 是一个 $n \times 1$ 列向量，则

$$\frac{\partial \boldsymbol{y}'}{\partial \boldsymbol{x}} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_2} \\ \vdots & \vdots & & \vdots \\ \frac{\partial y_1}{\partial x_n} & \frac{\partial y_2}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{pmatrix}$$

是一个 $n \times m$ 矩阵。

求导法则

- **列向量对矩阵求导**：若 $\boldsymbol{y} = (y_1, y_2, \cdots, y_m)'$ 是一个 $m \times 1$ 列向量，而 $\boldsymbol{X}$ 是一个 $p \times q$ 列向量，则

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{X}} = \begin{pmatrix} \frac{\partial y_1}{\partial \boldsymbol{X}} \\ \frac{\partial y_2}{\partial \boldsymbol{X}} \\ \vdots \\ \frac{\partial y_m}{\partial \boldsymbol{X}} \end{pmatrix}.$$

向量微商

常用符号及含义

设 $\boldsymbol{x}$ 是一个 $n \times 1$ 向量变量。

- 关于标量的微商：对于任意一个 $n \times 1$ 的非零常数向量 $\boldsymbol{a}$,
$$\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{a}'\boldsymbol{x}) = \frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}'\boldsymbol{a}) = \boldsymbol{a}.$$
- 关于二次型的微商：对于任意一个 $n \times n$ 的非零常数矩阵 $\boldsymbol{A}$,
$$\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}) = (\boldsymbol{A} + \boldsymbol{A}')\boldsymbol{x}.$$
特别地，若 $\boldsymbol{A} = \boldsymbol{A}'$，则 $\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}) = 2\boldsymbol{A}\boldsymbol{x}$.

矩阵微商

常用符号及含义

设 $\boldsymbol{X}$ 是一个 $n \times n$ 矩阵变量。

- 关于标量的微商：对于任意两个 $n \times 1$ 的非零常数矩阵 $\boldsymbol{a}, \boldsymbol{b}$,

$$\frac{\partial}{\partial \boldsymbol{X}}(\boldsymbol{a}'\boldsymbol{X}\boldsymbol{b}) = \boldsymbol{a}\boldsymbol{b}'.$$

  特别地，若 $\boldsymbol{a} = \boldsymbol{b}$，则 $\frac{\partial}{\partial \boldsymbol{X}}(\boldsymbol{a}'\boldsymbol{X}\boldsymbol{a}) = \boldsymbol{a}\boldsymbol{a}'$.

- 关于行列式的微商：

$$\frac{\partial}{\partial \boldsymbol{X}}|\boldsymbol{X}| = |\boldsymbol{X}|(\boldsymbol{X}^{-1})' = |\boldsymbol{X}'|(\boldsymbol{X}')^{-1}.$$

随机变量 vs 随机向量

- p 维随机向量（random vector）

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix},$$

其中，第 i 个分量 x_i 均是一个随机变量（random variable）；

- 当 $p = 1$ 时，$\boldsymbol{x} = x$ 是一个**标量**的随机变量（scalar random variable）。

随机向量

高斯分布随机向量

- 如果随机向量 $\boldsymbol{x} \sim N_p(\boldsymbol{\mu}, \Sigma)$，那么 $\boldsymbol{x}$ 的 p.d.f. 为

$$p(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\},$$

- 该分布为多维高斯分布；
- $\boldsymbol{\mu}$ 是一个 p 维向量，**可证明** $\boldsymbol{\mu} = E(\boldsymbol{x})$；
- Σ 是一个 $p \times p$ 正定实对称矩阵，**可证明** $\Sigma = \mathrm{Cov}(\boldsymbol{x})$。

证明： $\boldsymbol{\mu} = E(\boldsymbol{x})$

$$
\begin{aligned}
E(\boldsymbol{x}) &= \int_{R^p} \boldsymbol{x} p(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma) \mathrm{d}\boldsymbol{x} \\
&= \int_{R^p} \boldsymbol{x} (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right\} \mathrm{d}\boldsymbol{x} \\
&= \int_{R^p} (\boldsymbol{y} + \boldsymbol{\mu}) (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2} \boldsymbol{y}' \Sigma^{-1} \boldsymbol{y} \right\} \mathrm{d}\boldsymbol{y} \\
&= \int_{R^p} \boldsymbol{y} (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2} \boldsymbol{y}' \Sigma^{-1} \boldsymbol{y} \right\} \mathrm{d}\boldsymbol{y} \\
&\qquad + \boldsymbol{\mu} \int_{R^p} (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2} \boldsymbol{y}' \Sigma^{-1} \boldsymbol{y} \right\} \mathrm{d}\boldsymbol{y} \\
&= \boldsymbol{\mu}
\end{aligned}
$$

证明：$\Sigma = \mathrm{Cov}(\boldsymbol{x})$

因为 $\mathrm{Cov}(\boldsymbol{x}) = E(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})'$，所以，

$$
\begin{aligned}
\mathrm{Cov}(\boldsymbol{x}) &= \int_{R^p} (\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})' p(\boldsymbol{x};\boldsymbol{\mu},\Sigma)\mathrm{d}\boldsymbol{x} \\
&= \int_{R^p} (\boldsymbol{y}+\boldsymbol{\mu})(\boldsymbol{y}+\boldsymbol{\mu})'(2\pi)^{-p/2}\left|\Sigma\right|^{-1/2}\exp\left\{-\frac{1}{2}\boldsymbol{y}'\Sigma^{-1}\boldsymbol{y}\right\}\mathrm{d}\boldsymbol{y} \\
&= \int_{R^p} \boldsymbol{y}\boldsymbol{y}'(2\pi)^{-p/2}\left|\Sigma\right|^{-1/2}\exp\left\{-\frac{1}{2}\boldsymbol{y}'\Sigma^{-1}\boldsymbol{y}\right\}\mathrm{d}\boldsymbol{y} \\
&= \int_{R^p} \Sigma^{1/2}\boldsymbol{z}\boldsymbol{z}'\Sigma^{1/2}(2\pi)^{-p/2}\exp\left\{-\frac{1}{2}\boldsymbol{z}'\boldsymbol{z}\right\}\mathrm{d}\boldsymbol{z} \\
&= \Sigma^{1/2}\cdot\int_{R^p} \boldsymbol{z}\boldsymbol{z}'(2\pi)^{-p/2}\exp\left\{-\frac{1}{2}\boldsymbol{z}'\boldsymbol{z}\right\}\mathrm{d}\boldsymbol{z}\cdot\Sigma^{1/2} \\
&= \Sigma
\end{aligned}
$$

证明：$\Sigma = \mathrm{Cov}(\boldsymbol{x})$ （续）

只需要讨论 $\int_{R^p} \boldsymbol{z}\boldsymbol{z}'(2\pi)^{-p/2}\exp\left\{-\frac{1}{2}\boldsymbol{z}'\boldsymbol{z}\right\}\mathrm{d}\boldsymbol{z} = \boldsymbol{I}$ 的两种情况：

- 如果 $i \neq j$，那么

$$
\begin{aligned}
& \int_{R^p} z_i z_j \cdot (2\pi)^{-p/2}\exp\left\{-\frac{1}{2}\boldsymbol{z}'\boldsymbol{z}\right\}\mathrm{d}\boldsymbol{z} \\
= & \int_{R^2} z_i z_j (2\pi)^{-1}\exp\left\{-\frac{1}{2}\left(z_i^2+z_j^2\right)\right\}\mathrm{d}z_i\mathrm{d}z_j \\
= & \int_{R} z_i (2\pi)^{-1/2}\exp\left\{-\frac{1}{2}z_i^2\right\}\mathrm{d}z_i \\
& \cdot \int_{R} z_j (2\pi)^{-1/2}\exp\left\{-\frac{1}{2}z_j^2\right\}\mathrm{d}z_j \\
= & \ 0.
\end{aligned}
$$

证明：$\Sigma = \mathrm{Cov}(\boldsymbol{x})$ （续）

- 如果 $i = j$，那么

$$\begin{aligned} & \int_{R^p} z_i^2 \cdot (2\pi)^{-p/2} \exp\left\{-\frac{1}{2}\boldsymbol{z}'\boldsymbol{z}\right\} \mathrm{d}\boldsymbol{z} \\ = & \int_{R} z_i^2 (2\pi)^{-1/2} \exp\left\{-\frac{1}{2} z_i^2\right\} \mathrm{d}z_i \\ = & \ 1. \end{aligned}$$

综上，

$$\int_{R^p} \boldsymbol{z}\boldsymbol{z}' (2\pi)^{-p/2} \exp\left\{-\frac{1}{2}\boldsymbol{z}'\boldsymbol{z}\right\} \mathrm{d}\boldsymbol{z} = \boldsymbol{I}$$

随机向量

高斯分布随机向量

- 如果随机向量 $\boldsymbol{x} \sim N_p(\boldsymbol{\mu}, \Sigma)$，那么 $\boldsymbol{x}$ 的 p.d.f. 为

$$p(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \right\},$$

- 特别地，当 $\Sigma = \sigma^2 \boldsymbol{I}_p$ （$\boldsymbol{I}_p$ 是单位矩阵）时，

$$|\Sigma| = (\sigma^2)^p \quad 且 \quad \Sigma^{-1} = \sigma^{-2}\boldsymbol{I}_p,$$

于是，$\boldsymbol{x}$ 的 p.d.f. 为

$$p(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma) = (2\pi\sigma^2)^{-p/2} \exp\left\{ -\frac{1}{2\sigma^2}(\boldsymbol{x}-\boldsymbol{\mu})'(\boldsymbol{x}-\boldsymbol{\mu}) \right\}.$$

期望与协方差的性质

设随机变量 $\boldsymbol{x}=(x_1,x_2,\cdots,x_p)'$ 和 $\boldsymbol{y}=(y_1,y_2,\cdots,y_p)'$。设有 $m\times n$ 维常数矩阵 $\boldsymbol{A}=\{a_{ij}\}_{m\times n}$，$m'\times n$ 维常数矩阵 $\boldsymbol{B}=\{b_{ij}\}_{m'\times n}$ 及 m 维的常数向量 $\boldsymbol{c}=(c_1,c_2,\cdots,c_m)'$，可以证明以下结论：

- $E(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{c})=\boldsymbol{A}E(\boldsymbol{x})+\boldsymbol{c}$;
- $\mathrm{Cov}(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{c})=\boldsymbol{A}\mathrm{Cov}(\boldsymbol{x})\boldsymbol{A}'$;
- $\mathrm{Cov}(\boldsymbol{A}\boldsymbol{x},\boldsymbol{B}\boldsymbol{y})=\boldsymbol{A}\mathrm{Cov}(\boldsymbol{x},\boldsymbol{y})\boldsymbol{B}'$.

期望与协方差的性质

设随机变量 $\boldsymbol{x} = (x_1, x_2, \cdots, x_p)'$。若 $E(\boldsymbol{x}) = \boldsymbol{\mu}$，而 $\mathrm{Var}(\boldsymbol{x}) = \boldsymbol{\Sigma}$。对于任意对称正定矩阵 $\boldsymbol{A} = \{a_{ij}\}_{p\times p}$，有

$$E(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu} + \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}).$$

概率不等式

Jensen 不等式

给定一个概率空间 $(\Omega, \mathcal{F}, P)$，x 是一个随机变量且 ϕ 是一个凸函数，则

$$\phi(E(x)) \leq E(\phi(x)).$$

特殊形式

- 现有实数 $x_1, x_2, \cdots, x_n$ 且权重分别为 $a_1, a_2, \cdots, a_n$，有

$$\phi\left(\frac{\sum_{i=1}^n a_i x_i}{\sum_{i=1}^n a_i}\right) \leq \frac{\sum_{i=1}^n a_i \phi(x_i)}{\sum_{i=1}^n a_i}.$$

- 令 $\phi(x) = x^2$，则 $(E(x))^2 \leq E(x^2)$.

概率不等式

Hoeffding 不等式

设 $x_1, x_2, \cdots, x_n$ 是相互独立的随机变量且 $P(a_i \le x_i \le b_i) = 1$。令

$$S_n = x_1 + x_2 + \cdots + x_n.$$

对于任意 $t > 0$，都

$$\begin{aligned} P(S_n - E(S_n) \ge t) &\le \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right) \\ P(|S_n - E(S_n)| \ge t) &\le 2\exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right) \end{aligned}$$

无约束优化问题

无约束的最优化问题本质上是

$$\min f(\boldsymbol{x}), \boldsymbol{x} \in R^n.$$

符号与定义

- 目标函数： $f(\boldsymbol{x}), \boldsymbol{x} \in R^n.$
- 最小值点： $\boldsymbol{x}^*$， 即

$$\boldsymbol{x}^* = \arg\min f(\boldsymbol{x}),$$

- 优化问题的根本目标： 得到 $\boldsymbol{x}^*$ 解析解或近似解。

问题： 得到 $\boldsymbol{x}^*$ 的解析解的通法是什么?

迭代算法

梯度下降中解的迭代式为

$$\boldsymbol{x}_1 = \boldsymbol{x}_0 + \eta \cdot \left(- \left. \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} \right|_{\boldsymbol{x}=\boldsymbol{x}_0} \right)$$

其中,

- $- \left. \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} \right|_{\boldsymbol{x}=\boldsymbol{x}_0}$ 是梯度的反方向,表示迭代时的方向;
- η 是学习率,表示迭代时的步长;

梯度下降的原理

- 一阶泰勒展开公式为

$$f(\boldsymbol{x}_1) \approx f(\boldsymbol{x}_0) + \left(\left. \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} \right|_{\boldsymbol{x}=\boldsymbol{x}_0} \right)' (\boldsymbol{x}_1 - \boldsymbol{x}_0)$$

- 令

$$\boldsymbol{x}_1 - \boldsymbol{x}_0 = l\boldsymbol{v}$$

其中，
  - $\boldsymbol{v}$ 是一个单位向量，表示 $\boldsymbol{x}_1 - \boldsymbol{x}_0$ 的方向；
  - $l > 0$ 是一个标量，表示 $\boldsymbol{x}_1 - \boldsymbol{x}_0$ 的长度；

梯度下降的原理（续）

要使得 $f(\boldsymbol{x}_1) < f(\boldsymbol{x}_0)$，我们需要

$$\left(\left. \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} \right|_{\boldsymbol{x}=\boldsymbol{x}_0} \right)' l\boldsymbol{v} < 0$$

令 $\left(\left. \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} \right|_{\boldsymbol{x}=\boldsymbol{x}_0} \right) = \boldsymbol{u}$。因为 l 是一个正常数，不影响符号，所以，

$$\boldsymbol{u}'\boldsymbol{v} < 0$$

同时，我们知道

$$\boldsymbol{u}'\boldsymbol{v} = \langle \boldsymbol{u}, \boldsymbol{v} \rangle = \|\boldsymbol{u}\| \|\boldsymbol{v}\| \cos(\boldsymbol{u}, \boldsymbol{v}).$$

梯度下降的原理（续）

当 $\cos(\boldsymbol{u},\boldsymbol{v})=-1$ 时，$\boldsymbol{u}'\boldsymbol{v}$ 最小，于是取

$$\boldsymbol{v}=-\frac{\boldsymbol{u}}{\|\boldsymbol{u}\|}=-\frac{\left.\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\boldsymbol{x}_0}}{\left\|\left.\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\boldsymbol{x}_0}\right\|}.$$

因此，将 $\boldsymbol{v}$ 代入 $\boldsymbol{x}_1-\boldsymbol{x}_0=l\boldsymbol{v}$ 后可得

$$\boldsymbol{x}_1=\boldsymbol{x}_0+\eta\cdot\left(-\left.\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\boldsymbol{x}_0}\right).$$

带约束的优化问题

带约束的最优化问题本质上是

$$\begin{aligned}\min\quad & f(\boldsymbol{x}), \boldsymbol{x} \in R^n,\\ \text{s.t.}\quad & h_i(\boldsymbol{x}) \le 0, i = 1, 2, \cdots, m;\\ & l_i(\boldsymbol{x}) = 0, i = 1, 2, \cdots, r.\end{aligned}$$

符号与定义

- 拉格朗日函数为

$$L(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} u_i h_i(\boldsymbol{x}) + \sum_{j=1}^{r} v_i l_i(\boldsymbol{x})$$

其中，$\boldsymbol{u} = (u_1, \cdots, u_m)' \in R^m, \boldsymbol{v} = (v_1, \cdots, v_r)' \in R^r$
且 $u_i \ge 0$.

带约束的优化问题

性质

对于 $u_i \geq 0, i = 1, 2, \cdots, m$ 且 $v_i, i = 1, 2, \cdots, r$，有

$$f(\boldsymbol{x}) \geq L(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v})$$

原因如下：

$$L(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} u_i \underbrace{h_i(\boldsymbol{x})}_{\leq 0} + \sum_{j=1}^{r} v_j \underbrace{l_i(\boldsymbol{x})}_{=0} \leq f(\boldsymbol{x})$$

带约束的优化问题

令 $\mathcal{X}$ 是原始可行集。假设 f^* 是带约束的优化问题

$$\begin{aligned} \min \quad & f(\boldsymbol{x}), \boldsymbol{x} \in R^n, \\ \text{s.t.} \quad & h_i(\boldsymbol{x}) \leq 0, i = 1, 2, \cdots, m; \\ & l_i(\boldsymbol{x}) = 0, i = 1, 2, \cdots, r \end{aligned}$$

的原始最优值。于是有

$$f^* \geq \min_{\boldsymbol{x} \in \mathcal{X}} L(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}) \geq \min_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}) := g(\boldsymbol{u}, \boldsymbol{v})$$

我们称 $g(\boldsymbol{u}, \boldsymbol{v})$ 为拉格朗日对偶函数。

带约束的优化问题

给定原始优化问题

$$\begin{aligned} \min \quad & f(\boldsymbol{x}), \boldsymbol{x} \in R^n, \\ \text{s.t.} \quad & h_i(\boldsymbol{x}) \leq 0, i = 1, 2, \cdots, m; \\ & l_i(\boldsymbol{x}) = 0, i = 1, 2, \cdots, r. \end{aligned}$$

对偶函数 $g(\boldsymbol{u}, \boldsymbol{v})$ 满足：对于任意 $\boldsymbol{u} \geq 0$ 和 $\boldsymbol{v}$，有

$$f^* \geq g(\boldsymbol{u}, \boldsymbol{v}).$$

这样可以有**最优下界**：$\max_{\boldsymbol{u},\boldsymbol{v}} g(\boldsymbol{u}, \boldsymbol{v})$。
这自然导出了拉格朗日对偶问题为

$$\begin{aligned} \max_{\boldsymbol{u},\boldsymbol{v}} \quad & g(\boldsymbol{u}, \boldsymbol{v}) \\ \text{s.t.} \quad & \boldsymbol{u} \geq 0. \end{aligned}$$

带约束的优化问题

重要性质

- **弱对偶性**：如果对偶最优值为 g^*，那么
$$f^* \geq g^*.$$
注意到：这个性质总是成立的（即使原始问题是非凸的）。
- 对偶问题是一个凸优化问题（即使原始问题是非凸的）。
- **强对偶性**：如果对偶最优值为 g^*，那么
$$f^* = g^*.$$
注意到：这个性质成立时是**有条件的**。（Slater's 条件，KKT 条件）

带约束的优化问题

对偶问题的用途

- 在强对偶条件下，给定对偶最优值 $\boldsymbol{u}^*, \boldsymbol{v}^*$，原始最优解 $\boldsymbol{x}^*$ 也是
$$\min_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{u}^*, \boldsymbol{v}^*)$$
的解。

- 这表明：可以通过对偶问题的解来计算原始问题的解。

- 获取对偶问题的解，**可能更简单！**