

Part III-B: Deep Learning

Lecture by None

Note by THF

2025 年 2 月 4 日

目录

1 线性代数复习	1
1.1 特殊的矩阵和向量	3
1.2 特征分解	4
1.3 奇异值分解	6

Learn 1

01.28

1 线性代数复习

Definition. 张量：多维数组或多维立方体（二维为矩阵，一维为向量，零维为常量），用字体不同的粗体表示：**A**

对比：矩阵 **A**，张量 **A**

Definition. 向量：一般指列向量，转置后为横向量

Example. $\mathbf{x} = (1, 2, 3)^\top = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ ，转置后： $\mathbf{x}^\top = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$

Definition. 矩阵乘法：左矩阵的行遍历右矩阵的列

Example. 两个向量的点乘（内积）： $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^\top \mathbf{y}$ ，如下图：

同理，两个向量的外积： $\mathbf{x} \times \mathbf{y} = \mathbf{x} \mathbf{y}^\top$ ，如下图：

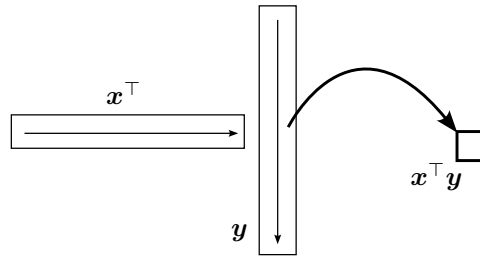


图 1: 点积

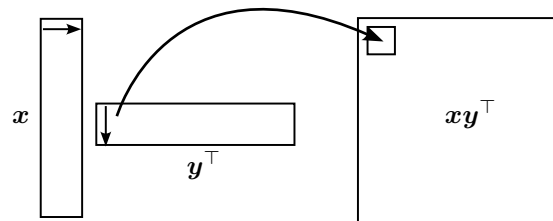


图 2: 外积

Definition. 广播: 把向量转置后加到矩阵的每一行

Example.

$$\begin{array}{ccc}
 \mathbf{A} & + \mathbf{b} & = \mathbf{C} \\
 \Rightarrow \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} & + \begin{bmatrix} b_1 & b_2 & b_3 \end{bmatrix}^\top & = \begin{bmatrix} a_{11} + b_1 & a_{12} + b_2 & a_{13} + b_3 \\ a_{21} + b_1 & a_{22} + b_2 & a_{23} + b_3 \end{bmatrix}
 \end{array}$$

Definition. 范数: 曼哈顿距离 \Rightarrow 欧氏距离 $\Rightarrow \dots$, L^p 范数形如:

$$\|\mathbf{x}\|_p = \left(\sum_i |x_i|^p \right)^{\frac{1}{p}}.$$

Example. 曼哈顿距离: $L^1: \|\mathbf{x}\|_1 = \sum_i |x_i|$, 欧几里得距离: $L^2: \|\mathbf{x}\|_2 = \sqrt{\sum_i x_i^2}$

Learn 2

01.30

Notation. L^2 范数可以写为 $\|\mathbf{x}\|_2$, 也可以简写为 $\|\mathbf{x}\|$

Notation. L^1 范数常用于区分 0 值和距离 0 很近的值, 使用 L^2 范数, 距离太近几乎没有区别

Definition. 最大范数: $p \rightarrow \infty$ 时影响范数的最大因素为最大值:

$$\|\mathbf{x}\|_{\infty} = \max_i |x_i|.$$

Example. Frobenius 范数: 使用范数衡量矩阵的大小, 类似于 L^2 范数的计算:

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} A_{i,j}^2}.$$

Example. 使用范数表示向量的内积:

$$\mathbf{x}^T \mathbf{y} = \|\mathbf{x}\| \cdot \|\mathbf{y}\| \cos \theta.$$

即:

$$\mathbf{x} \cdot \mathbf{y} = xy \cos \theta.$$

1.1 特殊的矩阵和向量

Example. 对角矩阵, 用 \mathbf{D} 表示, 除主对角线上的元素其他的元素均为零:

$$\mathbf{D} = \begin{bmatrix} a_{1,1} & 0 & \cdots & 0 \\ 0 & a_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{n,n} \end{bmatrix}.$$

Notation. 对角矩阵不一定是方阵, 非方阵超出方阵的部分全部为 0

Definition. 对角转换 diag : $\text{diag}(\mathbf{v})$ 将 \mathbf{v} 中的元素放入对角矩阵中

对角转换有以下性质:

- $\text{diag}(\mathbf{v})\mathbf{x} = \mathbf{v} \odot \mathbf{x}$
- $\text{diag}(\mathbf{v})^{-1} = \text{diag}\left(\left[\frac{1}{v_1} \quad \cdots \quad \frac{1}{v_n}\right]^T\right)$

Example. 对称矩阵: $\mathbf{A} = \mathbf{A}^T$, 常用于某些不依赖参数顺序的双参数函数, 交换 i, j 后结果不变

Example. 单位向量: 范数为 1 的向量, 即: $\|\mathbf{x}\|_n = 1$

Example. 向量的正交: $\mathbf{x}^T \mathbf{y} = 0$, 代表着 $\|\mathbf{x}\|_n$ 和 $\|\mathbf{y}\|_n$ 都不为零时 $\theta = \frac{\pi}{2}$
标准正交: $\theta = \frac{\pi}{2}$ 且 $\|\mathbf{x}\|_n = \|\mathbf{y}\|_n = 1$

Example. 正交矩阵：行向量和列向量都分别标准正交，有以下性质：

$$\mathbf{A}^\top \mathbf{A} = \mathbf{A} \mathbf{A}^\top = \mathbf{I} \Rightarrow \mathbf{A}^{-1} = \mathbf{A}^\top.$$

可得：正交矩阵求逆非常容易，可以用来求解线性方程组

1.2 特征分解

Definition. 特征向量和特征值：

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}.$$

则 λ 为 \mathbf{A} 的特征值， \mathbf{x} 为 \mathbf{A} 的一个（右）特征向量

$$\mathbf{x}^\top \mathbf{A} = \lambda \mathbf{x}^\top.$$

称 \mathbf{x} 为左特征向量

Definition. 特征分解：如果 \mathbf{A} 有 n 个线性无关的特征向量 $\{\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(n)}\}$ ，对应 n 个特征值 $\{\lambda_1, \dots, \lambda_n\}$ ，将特征向量（列向量）排列为一个矩阵 $\mathbf{V} = [\mathbf{v}^{(1)} \ \dots \ \mathbf{v}^{(n)}]$ ，并将特征值排列为一个向量 $\boldsymbol{\lambda} = [\lambda_1 \ \dots \ \lambda_n]^\top$ ，则：

$$\mathbf{A} = \mathbf{V} \text{diag}(\boldsymbol{\lambda}) \mathbf{V}^{-1}.$$

特征向量、值、分解的性质：

- \mathbf{A} 对 \mathbf{x} 变换相当于对 \mathbf{x} 缩放 λ 倍
- \mathbf{v} 是 \mathbf{A} 的特征向量，则 $c\mathbf{v}$ 也是，且特征值一样
- 实对称矩阵一定可以分解为实特征向量和实特征值，即 $\mathbf{A} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\top$

Notation. 实对称矩阵的分解：

$$\mathbf{A} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\top.$$

其中 \mathbf{Q} 为 \mathbf{A} 的特征向量 $\mathbf{v}^{(i)}$ 组成的正交矩阵， $\boldsymbol{\Lambda}$ 为对角矩阵，且 $\Lambda_{i,i} = \lambda_i$ 所对应的特征向量是 \mathbf{Q} 的第 i 个列向量 $\mathbf{Q}_{:,i}$

使用 \mathbf{A} 进行矩阵乘法可以看作是：将空间各自沿 $\mathbf{v}^{(i)}$ 延展 λ_i 倍

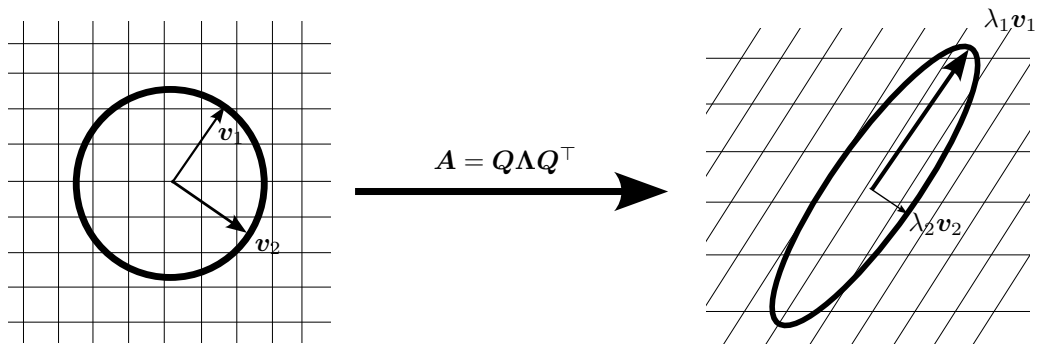


图 3: 实对称矩阵的乘法

Learn 3

01.31

Notation. $\mathbf{x}^\top \mathbf{A} \mathbf{x}$ 表示二次型或者二次方程 (\mathbf{x} 为二维向量), 如:

$$\begin{aligned} \mathbf{x} &= \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \mathbf{A} = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \\ \mathbf{x}^\top \mathbf{A} \mathbf{x} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \begin{bmatrix} x_1 a + x_2 b & x_1 b + x_2 c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= ax_1^2 + 2bx_1x_2 + cx_2^2. \end{aligned}$$

Definition. 所有特征值都是非负数的矩阵为**半正定**矩阵, 对于半正定矩阵有:

$$\forall \mathbf{x}, \mathbf{x}^\top \mathbf{A} \mathbf{x} \geq 0.$$

对于正定矩阵 ($A_{i,j} > 0$) 有:

$$\mathbf{x}^\top \mathbf{A} \mathbf{x} = 0 \Rightarrow \mathbf{x} = \mathbf{0}.$$

Notation. 当 \mathbf{x} 是 \mathbf{A} 的某个特征向量 \mathbf{v} 时, 由于:

$$\mathbf{A} \mathbf{v} = \lambda \mathbf{v}.$$

且 $\mathbf{x} = \mathbf{v} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^\top$, 二次型可以写为: $f = \mathbf{v}^\top \mathbf{A} \mathbf{v}$, 带入得:

$$\begin{aligned} f &= \mathbf{v}^\top \mathbf{A} \mathbf{v} \\ &= \mathbf{v}^\top \lambda \mathbf{v} \\ &= \lambda \mathbf{v}^\top \mathbf{v} \\ &= \lambda (x_1^2 + x_2^2). \end{aligned}$$

Learn 3

如果限定 $\|\mathbf{x}\|_2 = 1$, 即 $x_1^2 + x_2^2 = 1$, 则 $f = \lambda$, 即 $\max(f) = \max(\lambda), \min(f) = \min(\lambda)$

1.3 奇异值分解

除特征分解外, 还有一种分解叫做奇异值分解, 将一个实对称矩阵分解为**奇异向量**和**奇异值**

Notation. 特征分解: $\mathbf{A} = \mathbf{V} \text{diag}(\boldsymbol{\lambda}) \mathbf{V}^{-1}$

Definition. 奇异值分解 (SVD):

$$\mathbf{A} = \mathbf{U} \mathbf{D} \mathbf{V}^\top \quad \text{or} \quad \mathbf{A} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^\top.$$

Notation. 每个实数矩阵都有一个奇异值分解

Learn 4

02.01

Notation. 奇异值分解出来的部分:

- \mathbf{U} : $m \times m$ 的方阵, 为正交矩阵, 其列向量为**左奇异向量**
- \mathbf{D} : $m \times n$ 的矩阵, 为对角矩阵, 其对角线值为**奇异值**
- \mathbf{V} : $n \times n$ 的方阵, 为正交矩阵, 其列向量为**右奇异向量**