

# VIP Refresher: Linear Algebra and Calculus

Afshine AMIDI and Shervine AMIDI

October 13, 2018

翻译: 朱小虎

## 通用符号

□ **向量** – 我们记为一个  $n$  维的向量, 其中  $x_i \in \mathbb{R}$  是第  $i$  维的元素:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$$

□ **矩阵** – 我们记  $A \in \mathbb{R}^{m \times n}$  为一个  $m$  行  $n$  列的矩阵, 其中  $A_{i,j} \in \mathbb{R}$  是第  $i$  行  $j$  列的元素:

$$A = \begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{pmatrix} \in \mathbb{R}^{m \times n}$$

注意: 如上定义的向量  $x$  可以被看做是一个  $n \times 1$  的矩阵, 常被称为一个列向量。

□ **单位矩阵** – 单位矩阵  $I \in \mathbb{R}^{n \times n}$  是一个方阵其对角线上均是1 其余位置均为0。

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

注: 对所有矩阵  $A \in \mathbb{R}^{n \times n}$ , 我们有  $A \times I = I \times A = A$ 。

□ **对角阵** – 对角阵  $D \in \mathbb{R}^{n \times n}$  是一个方阵其对角线上元素均是非零值, 其余位置均为0。

$$D = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & d_n \end{pmatrix}$$

注: 我们记  $D$  为  $\text{diag}(d_1, \dots, d_n)$ 。

## 矩阵运算

□ **向量-向量** – 存在两种类型的向量-向量乘积:

- 内积: 对  $x, y \in \mathbb{R}^n$ , 我们有:

$$x^T y = \sum_{i=1}^n x_i y_i \in \mathbb{R}$$

- 外积: 对  $x \in \mathbb{R}^m, y \in \mathbb{R}^n$ , 我们有:

$$xy^T = \begin{pmatrix} x_1 y_1 & \cdots & x_1 y_n \\ \vdots & & \vdots \\ x_m y_1 & \cdots & x_m y_n \end{pmatrix} \in \mathbb{R}^{m \times n}$$

□ **矩阵-向量** – 矩阵  $A \in \mathbb{R}^{m \times n}$  和向量  $x \in \mathbb{R}^n$  的乘积是一个大小为  $\mathbb{R}^m$  的向量, 满足:

$$Ax = \begin{pmatrix} a_{r,1}^T x \\ \vdots \\ a_{r,m}^T x \end{pmatrix} = \sum_{i=1}^n a_{c,i} x_i \in \mathbb{R}^m$$

其中  $a_{r,i}^T$  是行向量,  $a_{c,j}$  是  $A$  的列向量,  $x_i$  是  $x$  的元素。

□ **矩阵-矩阵** – 矩阵  $A \in \mathbb{R}^{m \times n}$  和  $B \in \mathbb{R}^{n \times p}$  的乘积是一个大小为  $\mathbb{R}^{m \times p}$  的矩阵, 满足:

$$AB = \begin{pmatrix} a_{r,1}^T b_{c,1} & \cdots & a_{r,1}^T b_{c,p} \\ \vdots & & \vdots \\ a_{r,m}^T b_{c,1} & \cdots & a_{r,m}^T b_{c,p} \end{pmatrix} = \sum_{i=1}^n a_{c,i} b_{r,i}^T \in \mathbb{R}^{m \times p}$$

其中  $a_{r,i}^T, b_{r,i}^T$  是行向量,  $a_{c,j}, b_{c,j}$  分别是  $A$  和  $B$  的列向量

□ **转置** – 矩阵  $A \in \mathbb{R}^{m \times n}$  的转置, 记作  $A^T$ , 是其中元素的翻转

$$\forall i, j, \quad A_{i,j}^T = A_{j,i}$$

注: 对矩阵  $A, B$ , 我们有  $(AB)^T = B^T A^T$

□ **逆** – 可逆方阵  $A$  的逆记作  $A^{-1}$  和唯一满足下列要求的矩阵:

$$AA^{-1} = A^{-1}A = I$$

注: 不是所有方阵都是可逆的。同样, 对矩阵  $A, B$ , 我们有  $(AB)^{-1} = B^{-1}A^{-1}$

□ **迹** – 方阵  $A$  的迹, 记作  $\text{tr}(A)$ , 是对角线元素的和:

$$\text{tr}(A) = \sum_{i=1}^n A_{i,i}$$

注: 对矩阵  $A, B$ , 我们有  $\text{tr}(A^T) = \text{tr}(A)$  和  $\text{tr}(AB) = \text{tr}(BA)$

□ **行列式** – 方阵的行列式, 记作  $|A|$  或者  $\det(A)$  采用去掉第  $i$  行  $j$  列的矩阵  $A_{\setminus i, \setminus j}$  递归表达为如下形式:

$$\det(A) = |A| = \sum_{j=1}^n (-1)^{i+j} A_{i,j} |A_{\setminus i, \setminus j}|$$

注:  $A$  可逆当且仅当  $|A| \neq 0$ 。同样, 有  $|AB| = |A||B|$  和  $|A^T| = |A|$ 。

## 矩阵的性质

□ **对称分解** – 一个给定矩阵  $A$  可以用其对称和反对称部分进行表示:

$$A = \underbrace{\frac{A + A^T}{2}}_{\text{对称}} + \underbrace{\frac{A - A^T}{2}}_{\text{反对称}}$$

□ **范数** – 一个范数是一个函数  $N : V \rightarrow [0, +\infty]$  其中  $V$  是一个向量空间, 满足对所有  $x, y \in V$ , 有:

- $N(x + y) \leq N(x) + N(y)$
- 对一个标量  $a$ , 有  $N(ax) = |a|N(x)$
- 若  $N(x) = 0$ , 则  $x = 0$

对  $x \in V$ , 最常用的范数列在下表中:

范数	符号	定义	用例
曼哈顿, $L^1$	$\ x\ _1$	$\sum_{i=1}^n  x_i $	LASSO
欧几里德, $L^2$	$\ x\ _2$	$\sqrt{\sum_{i=1}^n x_i^2}$	Ridge
$p$ -范数, $L^p$	$\ x\ _p$	$\left(\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}}$	赫尔德不等式
无穷, $L^\infty$	$\ x\ _\infty$	$\max_i  x_i $	一致收

□ **线性相关** – 向量集合被称作线性相关的当其中一个向量可以被定义为其线性组合。

注: 若无向量可以按照此法表示, 则这些向量被称为线性无关。

□ **矩阵的秩** – 给定矩阵  $A$  的秩记作  $\text{rank}(A)$  是由列向量生成的向量空间的维度。这等于  $A$  的线性无关列向量的最大数目。

□ **半正定矩阵** – 矩阵  $A \in \mathbb{R}^{n \times n}$  是半正定矩阵 (PSD), 记作  $A \succeq 0$ , 当我们有:

$$A = A^T \quad \text{和} \quad \forall x \in \mathbb{R}^n, \quad x^T A x \geq 0$$

注: 类似地, 矩阵  $A$  被称作正定, 记作  $A \succ 0$ , 当它是一个 PSD 矩阵且满足所有非零向量  $x$ ,  $x^T A x > 0$ 。

□ **特征值, 特征向量** – 给定矩阵  $A \in \mathbb{R}^{n \times n}$ ,  $\lambda$  被称作  $A$  的一个特征值当存在一个向量  $z \in \mathbb{R}^n \setminus \{0\}$  称作特征向量, 满足:

$$Az = \lambda z$$

□ **谱定理** – 令  $A \in \mathbb{R}^{n \times n}$ , 若  $A$  是对称的, 则  $A$  可以被一个实正交矩阵  $U \in \mathbb{R}^{n \times n}$  对角化。记  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ , 我们有:

$$\exists \Lambda \text{ 对角阵, } A = U \Lambda U^T$$

□ **奇异值分解** – 对一个给定矩阵  $A$ , 其维度为  $m \times n$ , 奇异值分解 (SVD) 是一个因子分解技巧, 能保证存在酉矩阵  $U$   $m \times m$ , 对角阵  $\Sigma$   $m \times n$  和酉矩阵  $V$   $n \times n$ , 满足:

$$A = U \Sigma V^T$$

## 矩阵的微积分

□ **梯度** – 令  $f : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$  一个函数  $A \in \mathbb{R}^{m \times n}$  一个矩阵。  $f$  关于  $A$  的梯度是一个  $m \times n$  的矩阵, 记作  $\nabla_A f(A)$ , 满足:

$$\left(\nabla_A f(A)\right)_{i,j} = \frac{\partial f(A)}{\partial A_{i,j}}$$

注:  $f$  的梯度仅当  $f$  是返回一个标量的函数时有定义。

□ **Hessian** – 令  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  一个函数,  $x \in \mathbb{R}^n$  一个向量。  $f$  的关于  $x$  的 Hessian 是一个  $n \times n$  的对称阵, 记作  $\nabla_x^2 f(x)$ , 满足:

$$\left(\nabla_x^2 f(x)\right)_{i,j} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$$

注:  $f$  的 Hessian 仅当  $f$  是一个返回标量的函数时有定义。

□ **梯度运算** – 对矩阵  $A, B, C$ , 下列梯度性质值得记住:

$$\nabla_A \text{tr}(AB) = B^T$$

$$\nabla_A^T f(A) = (\nabla_A f(A))^T$$

$$\nabla_A \text{tr}(ABA^T C) = CAB + C^T A B^T$$

$$\nabla_A |A| = |A| (A^{-1})^T$$