

Repaso VIP: Álgebra Lineal y Cálculo

Afshine AMIDI y Shervine AMIDI

6 de octubre de 2018

Traducido por Fernando González-Herrera. Revisado por Fernando Diaz, Gustavo Velasco-Hernández y Juan P. Chavat.

Notaciones Generales

□ **Vector** – Sea $x \in \mathbb{R}^n$ un vector con n entradas, donde $x_i \in \mathbb{R}$ es la i -ésima entrada:

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

□ **Matriz** – Sea $A \in \mathbb{R}^{m \times n}$ una matriz con m filas y n columnas; donde $A_{i,j} \in \mathbb{R}$ es el valor ubicado en la i -ésima fila y la j -ésima columna:

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

Observación: el vector x definido arriba puede ser visto como una matriz $n \times 1$ y es particularmente llamado vector-columna.

□ **Matriz identidad** – La matriz identidad $I \in \mathbb{R}^{n \times n}$ es una matriz cuadrada con valores 1 en su diagonal y ceros en el resto:

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

Observación: para todas las matrices $A \in \mathbb{R}^{n \times n}$, se cumple que $A \times I = I \times A = A$.

□ **Matriz diagonal** – Una matriz diagonal $D \in \mathbb{R}^{n \times n}$ es una matriz cuadrada con valores diferentes de cero en su diagonal y cero en el resto:

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

Observación: también se denota D como $\text{diag}(d_1, \dots, d_n)$.

Operaciones de matrices

□ **Vector-vector** – Hay dos tipos de multiplicaciones vector-vector:

- Producto interno: para $x, y \in \mathbb{R}^n$, se tiene que:

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

- Producto diádico: para $x \in \mathbb{R}^m, y \in \mathbb{R}^n$, se tiene que:

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

□ **Matriz-vector** – El producto de la matriz $A \in \mathbb{R}^{m \times n}$ y el vector $x \in \mathbb{R}^n$, es un vector de tamaño \mathbb{R}^m ; tal que:

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

donde $a_{r,i}^T$ son los vectores fila y $a_{c,j}$ son los vectores columna de A , y x_i son las entradas de x .

□ **Matriz-matriz** – El producto de las matrices $A \in \mathbb{R}^{m \times n}$ y $B \in \mathbb{R}^{n \times p}$ es una matriz de tamaño $\mathbb{R}^{m \times p}$, tal que:

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

donde $a_{r,i}^T, b_{r,i}^T$ son los vectores fila y $a_{c,j}, b_{c,j}$ son los vectores columna de A y B respectivamente.

□ **Transpuesta** – La transpuesta de una matriz $A \in \mathbb{R}^{m \times n}$, denotada A^T , es tal que sus entradas se intercambian de la siguiente forma:

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

Observación: dadas las matrices A, B , se cumple que $(AB)^T = B^T A^T$.

□ **Inversa** – La inversa de una matriz cuadrada invertible A es denotada como A^{-1} y es la única matriz tal que:

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

Observación: no todas las matrices cuadradas son invertibles. Además, para las matrices A, B , se cumple que $(AB)^{-1} = B^{-1}A^{-1}$.

□ **Traza** – La traza de una matriz cuadrada A , denotada $\text{tr}(A)$, es la suma de los elementos en su diagonal:

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

Observación: dadas las matrices A, B , se cumple que $\text{tr}(A^T) = \text{tr}(A)$ y $\text{tr}(AB) = \text{tr}(BA)$.

□ **Determinante** – El determinante de una matriz cuadrada $A \in \mathbb{R}^{n \times n}$, denotado $|A|$ o $\det(A)$, es expresado recursivamente en términos de $A_{\setminus i, \setminus j}$, que es la matriz A sin su i -ésima fila ni su j -ésima columna, de la siguiente forma:

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

Observación: A es invertible si y solo si $|A| \neq 0$. Además, $|AB| = |A||B|$ y $|A^T| = |A|$.

Propiedades de matrices

□ **Descomposición simétrica** – Una matriz dada A puede ser expresada en términos de sus partes simétricas y antisimétricas de la siguiente forma:

$$A = \underbrace{\frac{A + A^T}{2}}_{\text{Simétrica}} + \underbrace{\frac{A - A^T}{2}}_{\text{Antisimétricas}}$$

□ **Norma** – Una norma (o módulo) es una función $N : V \rightarrow [0, +\infty[$ donde V es un vector espacial tal que para todo $x, y \in V$, se cumple que:

- $N(x + y) \leq N(x) + N(y)$
- $N(ax) = |a|N(x)$ siendo a un escalar
- si $N(x) = 0$, entonces $x = 0$

Para $x \in V$, las normas comúnmente utilizadas se resumen en la siguiente tabla:

Norma	Notación	Definición	Caso de uso
Manhattan, L^1	$\ x\ _1$	$\sum_{i=1}^n x_i $	LASSO
Euclidean, L^2	$\ x\ _2$	$\sqrt{\sum_{i=1}^n x_i^2}$	Ridge
p -norma, L^p	$\ x\ _p$	$\left(\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}}$	Desigualdad de Hölder
Infinito, L^∞	$\ x\ _\infty$	$\max_i x_i $	Convergencia

□ **Dependencia lineal** – Un conjunto de vectores se dice que es linealmente dependiente si uno de los vectores del conjunto puede ser definido como una combinación lineal de los restantes.

Observación: si ningún vector puede ser escrito de esta forma, entonces se dice que los vectores son linealmente independientes.

□ **Rango matricial** – El rango de una matriz dada A es denotado $\text{rank}(A)$ y es la dimensión del espacio vectorial generado por sus columnas. Esto es equivalente al máximo número de columnas linealmente independientes de A .

□ **Matriz semi-definida positiva** – Una matriz $A \in \mathbb{R}^{n \times n}$ es semi-definida positiva (PSD), lo cual se denota como $A \succeq 0$, si se cumple que:

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

Observación: de igual forma, una matriz A se dice definida positiva, lo cual se denota con $A \succ 0$, si es una matriz PSD que satisface para todos los vectores x diferentes de cero, $x^T A x > 0$.

□ **Valor propio, vector propio** – Dada una matriz $A \in \mathbb{R}^{n \times n}$, λ se dice que es un valor propio (*eigenvalue*, en inglés) de A si existe un vector $z \in \mathbb{R}^n \setminus \{0\}$, llamado vector propio (*eigenvector*, en inglés), tal que:

$$Az = \lambda z$$

□ **Teorema espectral** – Sea $A \in \mathbb{R}^{n \times n}$. Si A es simétrica, entonces A es diagonalizable a través de una matriz real ortogonal $U \in \mathbb{R}^{n \times n}$. Denotando $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, se tiene que:

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

□ **Descomposición en valores singulares** – Para una matriz dada A de dimensiones $m \times n$, la descomposición en valores singulares (SVD, por sus siglas en inglés de *Singular-Value Decomposition*) es una técnica de factorización que garantiza la existencia de las matrices U $m \times m$ unitaria, Σ $m \times n$ diagonal y V $n \times n$ unitaria, tal que:

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

Cálculo de matrices

□ **Gradiente** – Sea $f : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ una función y $A \in \mathbb{R}^{m \times n}$ una matriz. El gradiente de f con respecto a A es una matriz de $m \times n$, denotada por $\nabla_A f(A)$, tal que:

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

Observación: el gradiente de f se define únicamente cuando f es una función que retorna un escalar.

□ **Hessiana** – Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función y $x \in \mathbb{R}^n$ un vector. La hessiana de f con respecto a x es una matriz simétrica $n \times n$, denotada $\nabla_x^2 f(x)$, tal que:

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

Observación: la matriz hessiana de f se define únicamente cuando f es una función que devuelve un escalar.

□ **Operaciones de gradiente** – Dadas las matrices A, B, C , las siguientes propiedades de gradiente merecen ser tenidas en cuenta:

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

$$\nabla_{A^T} f(A) = (\nabla_A f(A))^T$$

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

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