Álgebra

Universidad Internacional de Valencia

Máster Universitario en Inteligencia Artificial

02MIAR | Matemáticas:

Matemáticas para la Inteligencia Artificial

Profesor:

Amílcar J. Pérez A.

universidadviu.com

De: Planeta Formación y Universidades

Definición

Un vector real v de dimensión n es una lista ordenada de n números reales:

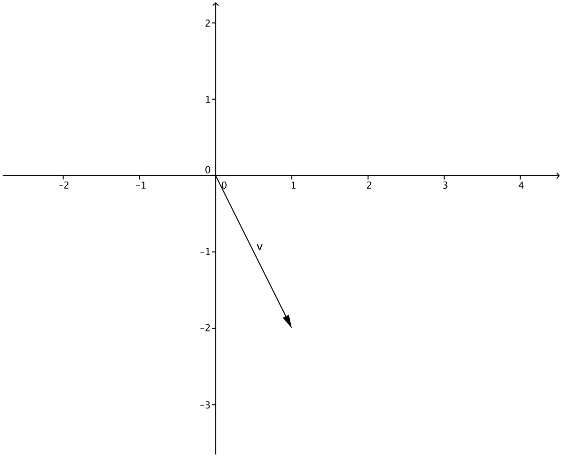
$$v = (v_1, v_2, \ldots, v_n), \quad v_i \in \mathbb{R}.$$

Se denota por $\mathbb{R}^n$ el conjunto formado por todos los vectores de dimensión n. Por tanto, podemos escribir $v \in \mathbb{R}^n$.

Ejemplos

1. $v = (1, -\frac{2}{3}) \in \mathbb{R}^2$.
2. $w = (\sqrt{2}, -1.2, \pi) \in \mathbb{R}^3$.

Los vectores en $\mathbb{R}^n$ se pueden representar gráficamente mediante un sistema de coordenadas n dimensional en caso de que $n \leq 3$.



Definición

Dados dos vectores $v, w \in \mathbb{R}^n$, $v = (v_1, v_2, \ldots, v_n)$, $w = (w_1, w_2, \ldots, w_n)$, *definimos su suma componente a componente:*

$$v + w = (v_1 + w_1, v_2 + w_2, \ldots, v_n + w_n).$$

Definición

Dado un número real λ *(un escalar) y un vector* $v \in \mathbb{R}^n$, $v = (v_1, v_2, \ldots, v_n)$, *definimos el producto de un escalar por un vector así:*

$$\lambda \cdot v = \lambda v = (\lambda v_1, \lambda v_2, \ldots, \lambda v_n).$$

Además de la longitud euclídea (noción usual longitud) es posible dar otras definiciones de la **norma** de un vector. Algunas normas importantes son:

- Norma euclídea (o norma 2):

$$\|v\| = \|v\|_2 = \sqrt{\sum_{i=1}^{n} v_i^2} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

- Norma 1:

$$\|v\|_1 = \sum_{i=1}^{n} |v_i| = |v_1| + |v_2| + \cdots + |v_n|.$$

- Norma del máximo:

$$\|v\|_\infty = \max_{1\leq i\leq n} |v_i|.$$

Ejemplo

Construimos un modelo computacional para predecir las temperaturas a partir de algunas variables de entrada (input). Supongamos que obtenemos un conjunto de n predicciones con el modelo considerado, almacenadas en un vector:

$$\hat{v} = (\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_n) \in \mathbb{R}^n.$$

Comparamos ahora el vector de predicciones, $\hat{v}$, con el vector de respuestas reales (observaciones), $v = (v_1, \ldots, v_n)$. Una forma de hacerlo es utilizar alguna norma $\|\cdot\|_*$ sobre la diferencia de los dos vectores:

$$E = \frac{1}{n}\|\hat{v} - v\|_*$$

Objetivo: encontrar un algoritmo que minimice E.

El producto escalar entre dos vectores $v, w \in \mathbb{R}^n$ se define como:

$$v \cdot w = \sum_{i=1}^{n} v_i w_i.$$

Ejemplo

Sean $v, w \in \mathbb{R}^3$*, con* $v = (2, -1, 0)$ *y* $w = (-3, -2, 1)$*. Entonces*

$$\begin{aligned} v \cdot w &= (2, -1, 0) \cdot (-3, -2, 1) \\ &= 2 \cdot (-3) + (-1) \cdot (-2) + 0 \cdot 1 = -6 + 2 + 0 = -4. \end{aligned}$$

Definición

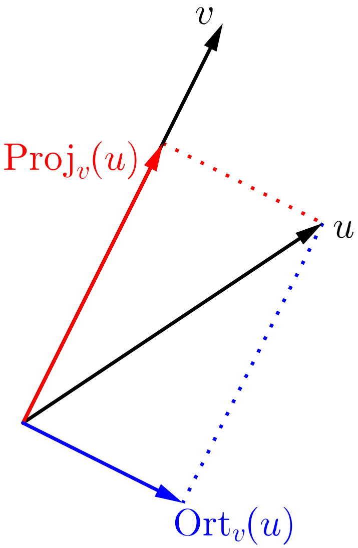
Sean $u, v \in \mathbb{R}^n$. Definimos la **proyección** *de u sobre v como*

$$\mathrm{Proj}_v(u) = \frac{u \cdot v}{v \cdot v} v = \frac{u \cdot v}{\|v\|^2} v$$

Definición

Sean $u, v \in \mathbb{R}^n$. Definimos la **ortogonal** *de u sobre v como*

$$\mathrm{Ort}_v(u) = u - \mathrm{Proj}_v(u)$$



Definición

Diremos que dos vectores $u, v \in \mathbb{R}^n$ *son* **ortogonales** *si* $u \cdot v = 0$.

La definición de ortogonalidad entre dos vectores se corresponde con la noción geométrica de perpendicularidad.

Más concretamente,

$$u \cdot v = \|u\|\|v\| \cos(\widehat{uv}),$$

donde $\widehat{uv}$ es el ángulo que forman los vectores u y v.

De esta expresión se sigue que, siendo $\hat{v} = v/\|v\|$ el vector unitario en la dirección de v:

$$\mathrm{Proj}_v(u) = \|u\| \cos(\widehat{uv})\hat{v}$$

Propiedades

Para $u, v, w \in \mathbb{R}^n$ *y* $\lambda \in \mathbb{R}$*:*

- $u + v = v + u.$
- $u \cdot v = v \cdot u.$
- $u \cdot (v + w) = u \cdot v + u \cdot w.$
- $(\lambda u) \cdot v = \lambda(u \cdot v).$
- $v \cdot v = \|v\|^2.$
- $u \cdot v = 0$ *si y sólo si* u *y* v *son ortogonales.*
- $v \cdot u = v \cdot \mathrm{Proj}_v(u).$
- $\mathrm{Ort}_v(u) \cdot v = 0$

Definición

Dados $v_1, v_2, \ldots, v_k \in V$, *una* **combinación lineal** *de los* k *vectores anteriores es cualquier expresión de la forma*

$$\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_k v_k,$$

con $\lambda_j \in \mathbb{R}$ *para* $j = 1, 2, \ldots, k$

Definición

Diremos que $v_1, v_2, \ldots, v_k \in V$ *son* **linealmente independientes** *si se cumple que la ecuación*

$$\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_k v_k = 0$$

tiene como única solución $\lambda_1 = \lambda_2 = \cdots = \lambda_k = 0$. *En caso contrario, se dice que éstos vectores son* **linealmente dependientes**.

Definición

Una **matriz** *de tamaño $m \times n$ sobre $\mathbb{R}$ es un conjunto formado por $m \cdot n$ números reales, a_{ij}, $1 \leq i \leq m$, $1 \leq j \leq n$, distribuidos en m filas y n columnas de la siguiente forma:*

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

Denotamos por $\mathbb{R}^{m\times n}$ el conjunto de todas las matrices $m \times n$ sobre $\mathbb{R}$. Una matriz $A \in \mathbb{R}^{m\times n}$ es **cuadrada** *si $m = n$, i.e., si tiene el mismo número de filas que de columnas. $I_n = (\delta_{ij})$ es la matriz identidad $n \times n$.*

Sean $A, B \in \mathbb{R}^{m\times n}$. Podemos definir la **suma** y la **resta** $A \pm B$ como:

$$A \pm B = \begin{pmatrix} a_{11} \pm b_{11} & a_{12} \pm b_{12} & \cdots & a_{1n} \pm b_{1n} \\ a_{21} \pm b_{21} & a_{22} \pm b_{22} & \cdots & a_{2n} \pm b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} \pm b_{m1} & a_{m2} \pm b_{m2} & \cdots & a_{mn} \pm b_{mn} \end{pmatrix}$$

Asimismo, dado un escalar $\lambda \in \mathbb{R}$, definimos el **producto** $\lambda \cdot A$ como

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

Por último, si $A \in \mathbb{R}^{m\times\ell}$ y $B \in \mathbb{R}^{\ell\times n}$, entonces definimos el **producto matricial** $A \cdot B \in \mathbb{R}^{m\times n}$ como

$$A \cdot B = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mn} \end{pmatrix}$$

donde $c_{ij} = \sum_{k=1}^{\ell} a_{ik} b_{kj} = a_{i1} b_{1j} + a_{i2} b_{2j} + \cdots + a_{i\ell} b_{\ell j}$.

Nota

En general, el producto de matrices **NO** *es conmutativo. Por ejemplo:*

$$\begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ -3 & 4 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -6 & 7 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 \\ -3 & 4 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} -1 & 2 \\ -7 & 8 \end{pmatrix}$$

Definición

Sea $A \in \mathbb{R}^{n\times n}$*. Diremos que* A *es* **invertible** *o* **regular** *si existe* $B \in \mathbb{R}^{n\times n}$ *tal que* $A \cdot B = B \cdot A = I_n$*. La matriz* B *recibe el nombre de* **matriz inversa** *y se denota por* A^{-1}*.*

Definición

Sea $A \in \mathbb{R}^{n\times n}$ *una matriz cuadrada de la forma*

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

Definimos el **determinante** *de* A *de forma recurrente como sigue:*

- *Si* $n = 1$, $\det(A) = a_{11}$.
- *Para* $n > 1$:

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

donde $A_{1j} \in \mathbb{R}^{(n-1)\times(n-1)}$ *es la matriz resultante de eliminar en* A *la fila* 1 *y la columna* j.

Teorema

El determinante de una matriz A puede calcularse de cualquiera de las siguientes formas:

- *Desarrollo por la fila i, $1 \leq i \leq n$:*

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}).$$

- *Desarrollo por la columna j, $1 \leq j \leq n$:*

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}).$$

donde $A_{ij} \in \mathbb{R}^{(n-1)\times(n-1)}$ es la matriz resultante de eliminar en A la fila i y la columna j.

Teorema

Para todo par de matrices $A, B \in \mathbb{R}^{n\times n}$, *se cumple*

$$\det(A \cdot B) = \det(A) \cdot \det(B).$$

Teorema

$A \in \mathbb{R}^{n\times n}$ *es regular si y sólo si* $\det(A) \neq 0$.

Definición

El **rango** *de una matriz* $A \in \mathbb{R}^{m\times n}$, *denotado por* $\operatorname{rank}(A)$, *es el número máximo de* **vectores columnas** *de* A *linealmente independientes.*

Teorema

$A \in \mathbb{R}^{n\times n}$ *es regular si y sólo si* $\operatorname{rank}(A) = n$.

Definición

Sea $A \in \mathbb{R}^{m\times n}$*. La* **traspuesta** *de la matriz* A *se define como la matriz* $A^\top \in \mathbb{R}^{n\times m}$ *tal que* $A^\top(i,j) = A(j,i)$, $1 \leq i \leq n$, $1 \leq j \leq m$.

Teorema

Sea $A \in \mathbb{R}^{n\times n}$ *regular. Entonces la inversa de* A, A^{-1}, *viene dada por*

$$A^{-1} = \frac{1}{\det(A)} \cdot \operatorname{adj}(A),$$

donde $\operatorname{adj}(A)$ *es la* **matriz adjunta** *de* A*, dada por* $\operatorname{adj}(A) = C^\top$, *con* $C(i,j) = (-1)^{i+j}|A_{ij}|$, $1 \leq i,j \leq n$, *con* $|A_{ij}| = \det(A_{ij})$.

Definición

Diremos que una aplicación $T : \mathbb{R}^n \to \mathbb{R}^m$ *es una* **aplicación lineal** *si* $\forall v, u \in \mathbb{R}^n$ *y* $\forall \alpha, \beta \in \mathbb{R}$ *se cumple*

$$T(\alpha v + \beta u) = \alpha T(v) + \beta T(u).$$

Ejemplos

1. *La aplicación* $T : \mathbb{R}^3 \to \mathbb{R}^2$ *dada por* $T(x, y, z) = (x + y, 2z - x)$ *es lineal.*
2. *La aplicación* $T : \mathbb{R}^2 \to \mathbb{R}^2$ *dada por* $T(x, y) = (x, x^2 - y)$ *no es lineal.*
3. *La aplicación* $T : \mathbb{R}^2 \to \mathbb{R}^2$ *dada por* $T(x, y) = (x - y, x + 1)$ *no es lineal.*

Teorema

Toda aplicación lineal $T : \mathbb{R}^n \to \mathbb{R}^m$ *puede representarse matricialmente en forma única (en las bases estándar de* $\mathbb{R}^n$ *y* $\mathbb{R}^m$*). Concretamente, existe una única* **matriz asociada** $M_T = [T] \in \mathbb{R}^{m\times n}$ *tal que, para todo* $v \in \mathbb{R}^n$ *puede escribirse:*

$$T(v) = M_T v.$$

Ejemplo

Sea $f : \mathbb{R}^3 \to \mathbb{R}^4$ *dada por*

$$T(x,y,z) = \begin{pmatrix} x+y \\ y-z \\ 2x+y+z \\ 3z \end{pmatrix} \longrightarrow M_T = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \\ 2 & 1 & 1 \\ 0 & 0 & 3 \end{pmatrix}$$

NOTA. Sea $\mathcal{B}_n = \{e_1, e_2, \ldots, e_n\}$ la base canónica de $\mathbb{R}^n$, i.e., $e_i = (\delta_{ij})$. Y sea $T : \mathbb{R}^n \to \mathbb{R}^m$ una transformación lineal. Entonces

$$[T] = (T(e_1) \mid T(e_2) \mid \ldots \mid T(e_n)) \quad \text{donde los } T(e_j) \text{ son vectores columna}$$

Esto se ve considerando que todo vector $v \in \mathbb{R}^n$ se escribe como $v = c_1 e_1 + c_2 e_2 + \ldots + c_n e_n = (c_1, c_2, \ldots, c_n)^\top$ y que por la linealidad de T se tiene que

$$T(v) = c_1 T(e_1) + c_2 T(e_2) + \ldots + c_n T(e_n) = \left(\begin{array}{c|c|c|c} T(e_1) & T(e_2) & \cdots & T(e_n) \end{array} \right) \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$$

Teorema

Sea $S : \mathbb{R}^n \to \mathbb{R}^\ell$ *y* $T : \mathbb{R}^\ell \to \mathbb{R}^m$ *aplicaciones lineales, con* M_S *y* M_T *las matrices asociadas respectivas. Entonces* $T \circ S : \mathbb{R}^n \to \mathbb{R}^m$ *es una aplicación lineal, cuya matriz asociada viene dada por* $M_{T\circ S} = M_T M_S$.

Ejercicios

1. *Comprueba que el teorema se cumple para las dos ejemplos anteriores.*
2. *Probar que* $T \circ S : \mathbb{R}^n \to \mathbb{R}^m$ *es una aplicación lineal si* $T : \mathbb{R}^\ell \to \mathbb{R}^m$ *y* $S : \mathbb{R}^n \to \mathbb{R}^\ell$ *lo son.*

Notación: La matriz asociada a T, M_T también la denotamos por $[T]$.

Ejemplo

Un ejemplo clásico de aplicaciones lineales son las rotaciones. En el caso de $\mathbb{R}^2$ *vienen dadas por* $T_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ *definidas como:*

$$T_\theta(x,y) = \begin{pmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \end{pmatrix} \longrightarrow [T_\theta] = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

donde θ *representa el ángulo de rotación (en sentido antihorario). Por tanto, esta aplicación envía un vector* $(x,y) \in \mathbb{R}^2$ *a su rotación en* θ *radianes:* $\left(x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta\right)$. *Asimismo, y como ya sabemos, podemos expresar la transformación matricialmente:*

$$T_\theta(x,y) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Ejemplo

Por ejemplo, la rotación de 30° (o $\pi/6$ rad) viene dada por:

$$[T_{\pi/6}] = \begin{pmatrix} \cos(\pi/6) & -\sin(\pi/6) \\ \sin(\pi/6) & \cos(\pi/6) \end{pmatrix} = \begin{pmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{pmatrix}$$

Por otra parte, la rotación de 60° (o $\pi/3$ rad) viene dada por:

$$[T_{\pi/3}] = \begin{pmatrix} \cos(\pi/3) & -\sin(\pi/3) \\ \sin(\pi/3) & \cos(\pi/3) \end{pmatrix} = \begin{pmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{pmatrix}$$

Por tanto, si componemos sendas rotaciones queda:

$$[T_{\pi/3} \circ T_{\pi/6}] = \begin{pmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{pmatrix} \begin{pmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = [T_{\pi/2}]$$

Ejemplo

Cabe esperar que el resultado sea una rotación de 90° *(o* $\pi/2$ *rad) con respecto al original:*

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -y \\ x \end{pmatrix}$$

cumpliéndose $(x, y) \cdot (-y, x) = -xy + yx = 0$*, luego son ortogonales.*

Ejercicio

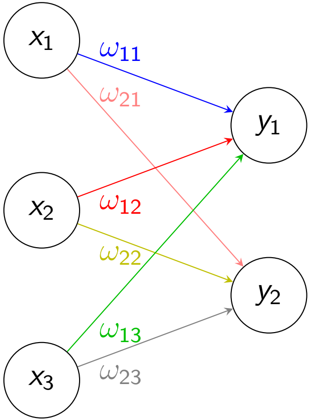
Utilizando las identidades trigonométricas que procedan, demuestra los siguientes enunciados:

(a) $\|T_\theta(v)\| = \|v\| \; \forall v \in \mathbb{R}^2, \forall \theta \in \mathbb{R}$.

(b) $T_\alpha \circ T_\beta = T_{\alpha+\beta} \; \forall \alpha, \beta \in \mathbb{R}$.

Motivación

Las matrices (y las aplicaciones lineales que representan) constituyen un elemento importante en la descripción de redes neuronales.



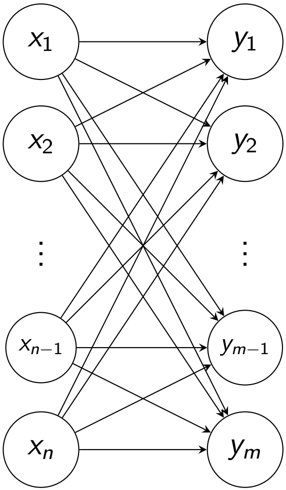
$$y_1 = f_1\left(\omega_{11}x_1 + \omega_{12}x_2 + \omega_{13}x_3 + b_1\right),$$
$$y_2 = f_2\left(\omega_{21}x_1 + \omega_{22}x_2 + \omega_{23}x_3 + b_2\right).$$

$$\Omega = \begin{pmatrix} \omega_{11} & \omega_{12} & \omega_{13} \\ \omega_{21} & \omega_{22} & \omega_{23} \end{pmatrix},\ X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

$$y_1 = f_1\left(\Omega(1,\cdot)X + b_1\right),$$
$$y_2 = f_2\left(\Omega(2,\cdot)X + b_2\right).$$

Motivación

En general, para el caso de información de una capa de n neuronas $(x_1, \ldots, x_n)$ que se transmite a otra capa de m neuronas $(y_1, \ldots, y_m)$:



$$y_i = f_i\left(\sum_{j=1}^{n} \omega_{ij} x_j + b_i\right), \; 1 \leq i \leq m.$$

$$\Omega = \begin{pmatrix} \omega_{11} & \omega_{12} & \cdots & \omega_{1n} \\ \omega_{21} & \omega_{22} & \cdots & \omega_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{m1} & \omega_{m2} & \cdots & \omega_{mn} \end{pmatrix},$$

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \therefore \; y_i = f_i\left(\Omega(i,\cdot)X + b_i\right), \; 1 \leq i \leq m.$$

¡Muchas gracias!

Universidad
Internacional
de Valencia

Contacto:

amilcar.perez@professor.universidadviu.com