PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

IMT3850: Fundamentos Matemáticos para Inteligencia Artificial
Clase 2

Profesor: Manuel A. Sánchez

Marzo 2023

Contenidos Clase 2:

- Matrices
- Multiplicación de Matrices
- Ecuaciones Lineales
- Clasificación Lineal
- Ejemplo: Dinámicas de Población
- Ejemplo: Dinámicas de Epidemias

Matrices

Matrices

Definición: Una matriz $A \in \mathbb{R}^{m\times n}$ es un arreglo rectangular de m-filas y n-columnas, formado por numeros reales.

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

$a_{ij} \in \mathbb{R}, 1 \leq i \leq m,\ 1 \leq j \leq n$ son las componentes de A

Ejemplos:

- Imagenes. Una imagen en blanco y negro de M por N pixels es representada por una matriz.
- Datos de precipitaciones. Precipitaciones en M lugares y N dias.
- Historial de compra. Historial de N clients y M productos.

Representación por columna

$x_1, x_2, \ldots, x_n \in \mathbb{R}^m$, entonces $X = [x_1|x_2|\ldots|x_n] \in \mathbb{R}^{m\times n}$

Definición:

- La matriz nula. $A = 0 \in \mathbb{R}^{m\times n}$, $a_{ij} = 0$, $1 \leq i \leq m$, $1 \leq j \leq n$
- La matriz identidad.

$$I \in \mathbb{R}^{n\times n},\ I_{ij} = \begin{cases} 1 & \text{si } i = j; \\ 0 & \text{si } i \neq j. \end{cases}$$

- Matriz diagonal.

$$A \in \mathbb{R}^{n\times n},\ A_{ij} = \begin{cases} a_{ii} & \text{si } i = j; \\ 0 & \text{si } i \neq j. \end{cases}$$

- Matriz triangular. $A \in \mathbb{R}^{m\times n}$

Superior: $a_{ij} = 0 \quad \text{si } i > j;$

Inferior: $a_{ij} = 0 \quad \text{si } i < j.$

- Matriz sparse. $A \in \mathbb{R}^{m\times n}$, $\text{nnz}(A) < mn$

Operaciones matriciales

- Transpuesta. Si $A \in \mathbb{R}^{m\times n}$, entonces su matriz transpuesta $A^\top \in \mathbb{R}^{n\times m}$ se define por

$$(A^\top)_{ji} = (A)_{ij}, \quad 1 \leq i \leq m, 1 \leq j \leq n.$$

Ademas, la matriz $A \in \mathbb{R}^{n\times n}$ se dice simétrica si $A = A^\top$.

- Adicion. Sean $A, B \in \mathbb{R}^{m\times n}$, entonces:

$$(A+B)_{ij} = (A)_{ij} + (B)_{ij}, \quad 1 \leq i \leq m, 1 \leq j \leq n$$

Propiedades de la adicion matricial:

» Conmutatividad. $A + B = B + A$

» Asociatividad. $(A+B)+C = A+(B+C)$

» Elemento identidad. $(A+0) = A$

» Transpuesta de la suma. $(A+B)^\top = A^\top + B^\top$

- Multiplicacion por escalar. Sea $A \in \mathbb{R}^{m\times n}$ y $\alpha \in \mathbb{R}$, entonces

$$(\alpha A)_{ij} = \alpha(A)_{ij}$$

Propiedas de la multiplicacion por escalar:

»Conmutatividad. $(\beta\alpha)A = \beta(\alpha A)$

»Asociatividad. $(\beta+\alpha)A = \beta A + \alpha A$

»Transpuesta. $(\alpha A)^\top = \alpha A^\top$

- Norma matricial. La norma de Frobenius (una de las normas matriciales) esta definida por

$$\begin{aligned} \|\cdot\|_F \ : \quad \mathbb{R}^{m\times n} &\rightarrow \quad \mathbb{R}_0^+ \\ A &\rightarrow \quad \|A\|_F := \sqrt{\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2} \end{aligned}$$

Muestre que:

- $\|A\|_F = \|A^\top\|_F$
- $\|A\|_F^2 = \|a_1\|^2 + \dots + \|a_n\|^2$

Producto matriz-vector

Sea $A \in \mathbb{R}^{m \times n}$ y $x \in \mathbb{R}^n$, entonces el product matriz vector se define

$$A = [a_1|\ldots|a_n], \quad Ax = y \in \mathbb{R}^m \quad \Longleftrightarrow \quad y_i = \sum_{j=1}^{n} a_{ij} x_j = a_1 x_1 + \ldots + a_n x_n$$

Ejemplos:

$$\begin{bmatrix} 1 & 2 \\ -1 & 0 \\ 4 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ -1 \\ 2 \end{bmatrix}$$

Ejemplos: Sean $D \in \mathbb{R}^{n-1\times n}, b \in \mathbb{R}^n$ definidas por

$$D_{ij} = \begin{cases} -1 & \text{si } i = j; \\ 1 & \text{si } j = i+1; \\ 0 & \text{en otro caso} \end{cases} \qquad b_i = f(\frac{i-1}{n-1}) = \sin(2\pi\frac{i-1}{n-1}),\ 1 \leq i \leq n$$

Entonces $$C = \frac{1}{n}Db$$

Observacion. Si $a, b \in \mathbb{R}^n$, el producto interior $a^\top b$ resulta de la multiplicacion matriz vector y la operacion transpuesta.

Multiplicación de Matrices

Multiplicacion de Matrices

Definición: Sea $A \in \mathbb{R}^{m\times p}$ y $B \in \mathbb{R}^{p\times n}$, entonces el producto matricial de A y B es un matriz $C \in \mathbb{R}^{m\times n}$ tal que

$$C_{ij} = \sum_{k=1}^{p} A_{ik}B_{kj}, \quad 1 \leq i \leq m,\, 1 \leq j \leq n$$

Ejemplo: Calcular el producto matricial de las siguientes matrices

$$A = \begin{bmatrix} 1 & -1 & 0 \\ 2 & 3 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & -1 \\ 0 & -1 \\ 1 & 0 \end{bmatrix}$$

Propiedades del producto matricial

Sean $A, \tilde{A} \in \mathbb{R}^{m\times p}, B \in \mathbb{R}^{p\times q}, C \in \mathbb{R}^{q\times n}, \alpha \in \mathbb{R}$. Entonces el producto matricial satisfice las siguiente propiedades:

- Asociatividad. $(AB)C = A(BC)$
- Asociatividad por escalar. $\alpha(AB) = (\alpha A)B$
- Distributividad con adición. $(A+\tilde{A})B = AB + \tilde{A}B$
- Transpuesta. $(AB)^\top = B^\top A^\top$

Observaciones.

1. Si $a \in \mathbb{R}^m$ y $b \in \mathbb{R}^n$ entonces el producto $ab^\top \in \mathbb{R}^{m\times n}$ se conoce como el producto exterior

$$ab^\top = \begin{bmatrix} a_1b_1 & a_1b_2 & \dots & a_1b_n \\ a_2b_1 & a_2b_2 & \dots & a_2b_n \\ \vdots & \vdots & \ddots & \vdots \\ a_mb_1 & a_mb_2 & \dots & a_mb_n \end{bmatrix}$$

2. Producto por matriz identidad $AI = IA = A$
3. Sean $A \in \mathbb{R}^{m\times p}, B \in \mathbb{R}^{p\times n}$, entonces en general

$$AB \neq BA$$

Ejemplos: Ilustraciones de producto matricial

1. Composición de funciones lineales.

$$\left.\begin{array}{rrcl} f: & \mathbb{R}^p & \longrightarrow & \mathbb{R}^m \\ & x & \longrightarrow & f(x) = Ax \\ g: & \mathbb{R}^n & \longrightarrow & \mathbb{R}^p \\ & x & \longrightarrow & g(x) = Bx \end{array}\right\} \quad \begin{array}{rrcl} h: & \mathbb{R}^n & \longrightarrow & \mathbb{R}^m \\ & x & \longrightarrow & h(x) = f(g(x)) = ABx \end{array}$$

2. Matriz de diferencias de orden 2. $\quad D_n \in \mathbb{R}^{n-1\times n}, D_{n-1} \in \mathbb{R}^{n-2\times n-1}$

Considere la ecuación diferencial: $\begin{cases} -u'' = \sin(2\pi x), & x \in (0,1) \\ u(0) = u(1) = 0 \end{cases}$

Calcule la solución del sistema lineal: $\dfrac{1}{(n-1)^2}(D_{n-1}D_n)x = b$

Factorización QR. Sea $A \in \mathbb{R}^{m \times n}$. Buscamos una matriz $Q \in \mathbb{R}^{m \times n}$ cuyos vectores columnas $q_1, q_2, \ldots, q_n$ sean vectores ortonormales y una matriz triangular superior $R \in \mathbb{R}^{n \times n}$, tales que

$$A = QR$$

Como podemos obtener la factorización QR de una matriz?

Matriz Inversa

Definición: Sea $A \in \mathbb{R}^{n\times n}$. Decimos que la matriz $X \in \mathbb{R}^{n\times n}$ es la matriz inver de A si

$$AX = XA = I$$

Denotamos la matriz inversa de A por A^{-1}.

Si la matriz inversa existe entonces decimos que A es invertible.

Ejemplos: Decida si las siguientes matrices son o no invertibles

$$1)\ A = \begin{bmatrix} 3 & 6 \\ 1 & 4 \end{bmatrix}$$

$$2)\ A = \begin{bmatrix} 3 & 6 \\ 2 & 4 \end{bmatrix}$$

Observación. Podemos extender el concepto de la matriz inversa a matrices no cuadradas $A \in \mathbb{R}^{m\times n}$

- Matriz Inversa por la izquierda si $XA = I \in \mathbb{R}^{n\times n}$
- Matriz Inversa por la derecha si $AX = I \in \mathbb{R}^{m\times m}$

Teorema. Sea $A \in \mathbb{R}^{m\times n}$ la inversa por la izquierda de A no existe si y solo si sus columnas son linealmente dependientes.

Demostración:

Suponga que las columnas de A son linealmente dependientes. Entonces existen coeficientes $x_1, ..., x_n$, no todos ceros, tales que:

$$a_1x_1 + ... + a_nx_n = 0 \quad \iff \quad Ax = 0; \quad A = [a_1|...|a_n],\ x = (x_1, ..., x_n)^\top$$

Suponga por contradiccion que existe la inversa por la izquierda C, entonces

$$0 = C(Ax) = CAx = Ix = x,$$

lo que implica que x es el vector nulo, lo que es una contradicción.

Propiedades de la inversa.

Sea $A, B \in \mathbb{R}^{n\times n}$ invertible y $\alpha \in \mathbb{R}$. Entonces

- $(A^{-1})^{-1} = A$
- $(\alpha A)^{-1} = (1/\alpha)A^{-1}$
- $(AB)^{-1} = B^{-1}A^{-1}$
- $(A^\top)^{-1} = (A^{-1})^\top = A^{-\top}$

Ecuaciones Lineales

Ecuaciones Lineales

Funciones vectoriales:
$$\begin{aligned} f \quad : \mathbb{R}^n &\longrightarrow \mathbb{R}^m \\ x &\longrightarrow f(x) = (f_1(x), f_2(x), \dots, f_m(x)) \end{aligned}$$

Ejemplo: Sea $A \in \mathbb{R}^{m\times n}$ y
$$\begin{aligned} f \quad : \mathbb{R}^n &\longrightarrow \mathbb{R}^m \\ x &\longrightarrow f(x) = Ax \end{aligned}$$

Observe que esta satisface, para $x, y \in \mathbb{R}^n$ y $\alpha, \beta \in \mathbb{R}$

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$$

Definición: Sea $A \in \mathbb{R}^{m\times n}$. Entonces la imagen de A se define por

$$ran(A) := \{y \in \mathbb{R}^m : \ y = Ax, \ \text{para } x \in \mathbb{R}^n\}$$

Observe que si $A = [a_1|...|a_n]$ entonces $ran(A) = \text{span}\{a_1, ..., a_n\}$

Además, definimos el rango de A como la dimension del espacio imagen, esto es

$$rank(A) = \dim(ran(A))$$

Definimos el espacio nulo o kernel de la matriz $A \in \mathbb{R}^{m\times n}$ por

$$Ker(A) = \{x \in \mathbb{R}^n : Ax = 0\}$$

Teorema. Si $A \in \mathbb{R}^{m\times n}$, entonces

$$\dim(Ker(A)) + rank(A) = n$$

Observación. Una matriz $A \in \mathbb{R}^{m\times n}$ se dice de rango deficiente si

$$rank(A) < \min\{m, n\}$$

Modelos de funciones lineales.

Aproximación de Taylor. Sea la función $f: \mathbb{R}^n \longrightarrow \mathbb{R}^m$ diferenciable. La expansión en serie de Taylor en $x = z \in \mathbb{R}^n$ de $f(x) = (f_1(x), ..., f_m(x))$

$$
\begin{aligned}
f_i(x) &\approx f_i(z) + \frac{\partial f_i}{\partial x_1}(z)(x_1 - z_1) + ... + \frac{\partial f_i}{\partial x_n}(z)(x_n - z_n) \\
f_i(x) &\approx f_i(z) + \nabla f_i^\top(z)(x - z)
\end{aligned}
$$

$$
\nabla f_i(z) = \begin{bmatrix} \partial f_i/\partial x_1 \\ \vdots \\ \partial f_i/\partial x_n \end{bmatrix}
$$

$$
f(x) = f(z) + \nabla f(z)(x - z)
$$

$$
\begin{aligned}
J = \nabla f(x) &= \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \cdots & \frac{\partial f_m}{\partial x_n}(x) \end{bmatrix} \\
&= \left[\frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_n} \right]
\end{aligned}
$$

Sistemas de ecuaciones lineales

Dados $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$

Hallar $x \in \mathbb{R}^n$ tal que

$$Ax = b \quad \Longleftrightarrow \quad \begin{cases} n\text{-incógnitas y } m\text{-ecuaciones} \\ A_{11}x_1 + \ldots + A_{1n}x_n = b_1 \\ A_{11}x_1 + \ldots + A_{1n}x_n = b_1 \\ \qquad \vdots \\ A_{11}x_1 + \ldots + A_{1n}x_n = b_1 \end{cases}$$

Definición: El Sistema de ecuaciones lineales se dice:

- Sobredeterminado si $m > n$
- Subdeterminado si $m < n$
- Cuadrado si $m = n$

Clasificación Lineal

Clasificación Lineal

Ejemplo: *Motivación*. Una librería tiene una lista de clientes $\{\boldsymbol{c}\}$ con cada $\boldsymbol{c} = (c_1, c_2, ..., c_n), n -$ features. Entonces se quiere determinar a que clientes recomendar un determinado libro L.

Hipótesis de linealidad. Asumimos que la decisión de recomendar o no el libro L a $\boldsymbol{c}$ depende linealmente de los features de $\boldsymbol{c}$.

Esto lo podemos representar como

$$w_1 c_1 + w_2 c_2 + ... + w_n c_n \geq w_0$$

si y solo si es recomendable ofrecer el libro L a C.

Reescribimos: $$w^\top C \geq w_0$$

Como determinar w ?

Buscamos en la base de datos ejemplos de clientes que compraron o no el libro.

Ejemplos positivos

$$\begin{array}{rcl} w_1 p_{11} + w_2 p_{12} + \ldots + w_n p_{1n} & \geq & w_0 \\ \vdots & \geq & \vdots \\ w_1 p_{m1} + w_2 p_{m2} + \ldots + w_n p_{mn} & \geq & w_0 \end{array}$$

Ejemplos negativos

$$\begin{array}{rcl} w_1 q_{11} + w_2 q_{12} + \ldots + w_n q_{1n} & < & w_0 \\ \vdots & < & \vdots \\ w_1 q_{r1} + w_2 q_{r2} + \ldots + w_n q_{rn} & < & w_0 \end{array}$$

El problema de "separar" un conjunto de datos utilizando hiperplanos se conoce se conoce como clasificación lineal.

Observación.

- Un conjunto de datos no es necesariamente separable linealmente
- Un conjunto de datos puede transformarse para que si pueda ser separable linealmente.
- Regresion lineal.

Ejemplos positivos

$$P \in \mathbb{R}^{m\times n}, \quad P\boldsymbol{w} \geq w_0 \quad \longrightarrow \quad \widehat{P}\boldsymbol{y} \geq 0$$

Ejemplos negativos

$$Q \in \mathbb{R}^{r\times n}, \quad Q\boldsymbol{w} < w_0 \quad \longrightarrow \quad \widehat{Q}\boldsymbol{y} < 0$$

Algoritmo: Perceptron

Input: e

Output: $y \in \mathbb{R}^n$

$y_0 = 0 \in \mathbb{R}^n$

while y no se solución :

Elija e mal clasificado por y

$$\text{Actualice } y = \eta * \begin{cases} (y + e) & \text{si } e \text{ es positivo} \\ (y - e) & \text{si } e \text{ es negativo} \end{cases}$$

Sistemas dinámicos lineales: Aplicación de producto matriz-vector

Sean los vectores $x_1, x_2, \ldots, x_m \in \mathbb{R}^n$, aquí el índice de los vectores indica los estados de una variable en tiempo.

Sistema dinámico lineal:

$$x_{t+1} = A_t x_t$$

Extensiones:

- + input

$$x_{t+1} = A_t x_t + B_t u_t + C_t$$

- Modelo de K-Markov

$$x_{t+1} = A_1 x_t + A_2 x_{t-1} + \ldots + A_K x_{t-K+1}, \quad t = K, K+1, \ldots$$

Ejemplo: **Dinámicas de población.**

- Describir la evolución de la distribución de edad en una población.

$$x_t \in \mathbb{R}^{100},\ (x_t)_i : \quad \text{número de personas con edad } i-1 \text{ en año } t.$$

Tenemos la tasa de natalidad y mortalidad $b \in \mathbb{R}^{100},\ d_t \in \mathbb{R}^{100}$

$(b)_i$: el número promedio de nacimientos por persona de edad $i-1$.

$(d)_i$: fracción de las personas de edad $i-1$ que morirán ese año.

Construimos el modelo:

$$\begin{aligned}(x_{t+1})_1 &= b^\top x_t \\ (x_{t+1})_{i+1} &= (1-d)(x_t)_i,\ i = 1, \ldots, 99\end{aligned}$$

$$A = \begin{bmatrix} b_1 & b_2 & \ldots & b_{99} & b_{100} \\ 1-d_1 & 0 & 0 & \ldots & 0 \\ 0 & 1-d_2 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \ldots & 0 & 1-d_{99} & 0 \end{bmatrix}$$

$$\boxed{x_{t+1} = Ax_t}$$

Predecir :

- población
- niños en edad escolar
- adultos en edad de jubilación

Ejemplo: **Dinámicas de epidemias.**

$$\text{Fracción de la población en: } \left.\begin{array}{r}\text{Susceptible}\\ \text{Infectado}\\ \text{Recuperado}\\ \text{Difuntos}\end{array}\right\} \quad x_t \in \mathbb{R}^4$$

Suponemos en nuestro modelo simple que cada día

- 5 % de la población es susceptible tendrá la enfermedad
- 1% de la población infectada fallecerá
- 10% de la población infectada se recuperara y quedara inmune
- 4% de la población infectada se recuperara pero no quedara inmune

$$x_{t+1} = \begin{bmatrix} 0.95 & 0.04 & 0 & 0 \\ 0.05 & 0.85 & 0 & 0 \\ 0 & 0.1 & 1 & 0 \\ 0 & 0.01 & 0 & 1 \end{bmatrix} x_t$$