Conceptos de Álgebra Lineal para ML

I302 - Aprendizaje Automático y Aprendizaje Profundo

Roberto Bunge

Universidad de San Andrés

y = A x

$y = Ax$ y es la proyección de x a través de la matriz A

$A \in R^{m \times n}$

$x \in R^n$

$y \in R^m$

$$y \in R^m \longleftarrow \boxed{A} \longleftarrow x \in R^n$$

$$A = [a_1 \dots a_n] = \begin{bmatrix} \tilde{a}_1^T \\ \vdots \\ \tilde{a}_n^T \end{bmatrix}$$

$$y = Ax = \sum_{i=1}^{n} a_i \cdot x_i = a_1 \cdot x_1 + a_2 x_2 \cdots + a_n \cdot x_n$$

(y es la combinación lineal de los columnas de A con los coeficientes x_i)

$$y_i = \tilde{a}_i^T \cdot x$$

(el elemento y_i es el producto interno entre las filas de A con x)

Imagen y Nulidad (de una matriz cualquiera)

Sea $A \in \mathbb{R}^{m \times n}$

Imagen: $\mathrm{Im}(A) = \{ y \in \mathbb{R}^m \mid \exists x \in \mathbb{R}^n,\ y = Ax \}$

= subespacio generado por las columnas de A

= todos los vectores $\in \mathbb{R}^m$ "alcanzables" por una proyección de A

Nulidad: $\mathrm{Nul}(A) = \{ x \in \mathbb{R}^n \mid 0 = A \cdot x \}$

= vectores ortogonales a las filas de A

= vectores que son "aplastados" por A (no tienen proyección)

A es "uno a uno" $\iff \mathrm{Nul}(A) = \{0\} \iff$ columnas de A son L.I $\Rightarrow |A^T A| \neq 0$

A es "onto" $\iff \mathrm{Im}(A) = \mathbb{R}^n \iff$ columnas de A son base L.I de $\mathbb{R}^m$

$\iff$ filas de A son L.I.

$\iff \mathrm{Nul}(A^T) = \{0\} \Rightarrow |AA^T| \neq 0$

Rango de una matriz (cualquier matriz)

Sea $A \in \mathbb{R}^{m \times n}$

$$\text{rango}(A) = \dim(\text{Im}(A)) = \dim(\text{Im}(A^T)) = \text{rango}(A^T)$$

$$\leq \min(m,n)$$

La matriz A es de "rango completo" $\iff$ $\text{rango}(A) = \min(m,n)$

Sea A de rango completo:

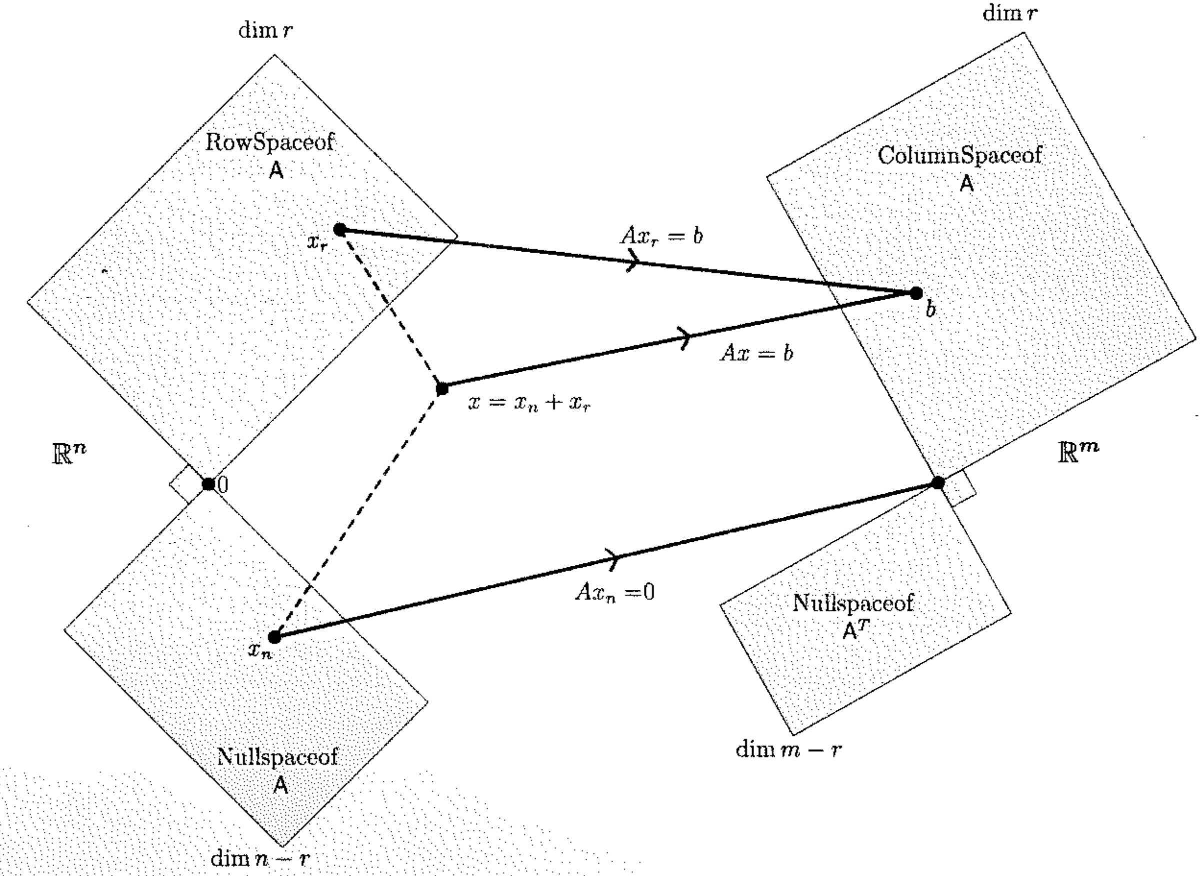
- si $m=n$ (cuadrada) $\Rightarrow$ A es invertible
- si $m>n$ ("delgada") $\Rightarrow$ columnas son independientes
- si $m<n$ ("ancha") $\Rightarrow$ filas son independientes

Teorema Fundamental del Álgebra Lineal

Sea A matriz m x n

$Nul(A) \perp Im(A^T)$

$Im(A) \perp Nul(A^T)$



* Tomado de https://mathworld.wolfram.com/FundamentalTheoremofLinearAlgebra.html

Inversa de una matriz cualquiera

matriz identidad $\in \mathbb{R}^{n\times n}$

Sea $A \in \mathbb{R}^{m\times n}$

si $\text{rango}(A) = n \Longrightarrow \exists$ inversa izquierda $B \in \mathbb{R}^{n\times m} \;/\; B.A = I_n$

si $\text{rango}(A) = m \Longrightarrow \exists$ inversa derecha $B \in \mathbb{R}^{m\times n} \;|\; A.B = I_m$

Inversa de una matriz cuadrada

Sea $A \in \mathbb{R}^{n \times n}$ (cuadrada)

es única

Si rango$(A) = n \iff |A| \neq 0 \iff \exists$ inversa (izquierda y derecha) $A^{-1} \mid A^{-1}A = I_n$, $AA^{-1} = I_n$

$\iff$ columnas y filas de A son LI (y base de $\mathbb{R}^n$)

$\iff Nul(A) = \{0\}$

$\iff Im(A) = \mathbb{R}^n$

$\Rightarrow |A^TA| \neq 0$ (A^TA es invertible)

$\Rightarrow |AA^T| \neq 0$ (AA^T es invertible)

Matriz Ortonormal

Sea $\{u_1, \dots, u_k\} \in \mathbb{R}^n$ un conjunto de vectores ortonormales $\iff$ (i) $u_i \perp u_j \quad i \neq j$

(ii) $\|u_i\| = 1$

Si definimos $U = [u_1 \cdots u_k] \iff U^T U = I$ (pero no necesariamente $UU^T = I$)

Sea $y = Ux \implies \|y\| = \|x\|$ (la transformación U preserva distancias)

$\langle Uz, U\bar{z} \rangle = \langle z, \bar{z} \rangle$ (la transformación U preserva ángulos)

Matriz Ortogonal

Sea $\{u_1, \dots, u_n\} \in \mathbb{R}^n$ un conjunto de n vectores ortonormales $\Longleftrightarrow$ es base ortonormal de $\mathbb{R}^n$

Si definimos $U = [u_1 \dots u_n]$ $\Longrightarrow$ U es una matriz "ortogonal"

$U \in \mathbb{R}^{n \times n}$

$\Longrightarrow U^T U = I$

$\Longrightarrow U U^T = I$

$\Longrightarrow U^{-1} = U^T$

Autovectores y autovalores (de matriz cuadrada)

Sea $A \in \mathbb{R}^n$

números complejos

Autovalor: Sea $\lambda \in \mathbb{C}$ / $|\lambda I - A| = 0 \iff \lambda$ es autovalor de A

Toda matriz cuadrada tiene n autovalores, con posibles repeticiones.
Para encontrar los autovalores de A debemos encontrar las raíces λ_i del "polinomio característico" que surge de $|\lambda I - A| = 0$

Autovector derecho: Sea $v \in \mathbb{C}^n$, $v \neq 0$, $\lambda \in \mathbb{C}$ / $Av = \lambda v \iff v$ es un "autovector derecho" de A y λ es el autovalor asociado a v

Autovector izquierdo: Sea $w \in \mathbb{C}^n$, $w \neq 0$, $\lambda \in \mathbb{C}^n$ / $w^T A = \lambda w^T \iff w$ es un "autovector izquierdo" de A y λ es el autovalor asociado a w

Diagonalización (de matriz cuadrada)

Sea $A \in \mathbb{R}^{n \times n}$

si $\exists \{v_1, \dots, v_n\}$ conjunto de autovectores derechos L.I. $\iff$ A es diagonalizable

si definimos $T = [v_1 \cdots v_n]$ $\iff A = T \Lambda T^{-1}$

$\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$ $\iff \Lambda = T^{-1} A T$

$\Rightarrow T^{-1} = \begin{bmatrix} w_1^T \\ \vdots \\ w_n^T \end{bmatrix}$ son autovectores izquierdos L.I.

No todas las matrices cuadradas son diagonalizables

AutoVec/Val y Diagonalización de Matrices Simétricas

Sea $A \in R^{n \times n}$ simétrica $\iff A = A^T \implies$ autovalores $\lambda(A) \in R$ (los autovalores son reales)

$\implies \exists \{q_1, \ldots, q_n\}$ conjunto de autovectores derechos e izquierdos de A

Si definimos $Q = [q_1 \cdots q_n] \implies A = Q \Lambda Q^{-1} = Q \Lambda Q^T$

$\Lambda = diag(\lambda_1, \ldots, \lambda_n) \qquad \implies \Lambda = Q^T A Q$

Q es matriz ortogonal $\implies Q^{-1} = Q^T$

$\implies Q^T Q = I$

$\implies Q Q^T = I$

Toda matriz simétrica es diagonalizable.

Descomposición en Valores Singulares (de cualquier matriz)

Sea $A \in \mathbb{R}^{m \times n}$ con rango$(A) = r$

$\Rightarrow \exists\ U \in \mathbb{R}^{m \times r}$, $U = [u_1 \cdots u_r]$, $U^T U = I$, u_i son vectores singulares derechos de A

$V \in \mathbb{R}^{n \times r}$, $V = [v_1 \cdots v_r]$, $V^T V = I$, v_i son vectores singulares izquierdos de A

$\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_r)$, $\sigma_i \in \mathbb{R}$ son valores singulares de A asociados a los vectores singulares u_i, v_i

$$A = U \Sigma V^T = \sum_{i=1}^{r} \sigma_i \, u_i v_i^T$$

$$A = {}_m\overset{r}{\boxed{U}}\ {}_r\overset{r}{\boxed{\Sigma}}\ {}_r\overset{n}{\boxed{V^T}}$$

$A^T A = V \Sigma^2 V^T \Rightarrow v_i$ autovectores de $A^T A$

$\Rightarrow \sigma_i^2 = \lambda_i(A^T A) \Rightarrow \sigma_i = \sqrt{\lambda_i(A^T A)}$ $\quad (\lambda_i(A^T A) = 0,\ i > r)$

$A A^T = U \Sigma^2 U^T \Rightarrow u_i$ autovectores de $A A^T$ $\quad (\lambda_i(A A^T) = 0,\ i > r)$

$\Rightarrow \sigma_i^2 = \lambda_i(A A^T) \Rightarrow \sigma_i = \sqrt{\lambda_i(A A^T)}$

Pseudo-inversa de una matriz cualquiera

Sea $A \in \mathbb{R}^{m \times n}$ de rango completo $\Leftrightarrow$ rango$(A) = r = \min(m,n)$

$\Rightarrow A^+ = V\Sigma^{-1}U^T$ es la pseudo-inversa de A

También conocida como la inversa de Moore-Penrose

Si $m > n$ (A es "delgada") $\Rightarrow A^+ = (A^TA)^{-1}A^T \Rightarrow A^+A = I$

Si $m < n$ (A es "ancha") $\Rightarrow A^+ = A^T(AA^T)^{-1} \Rightarrow AA^+ = I$

Si $m = n$ (A es cuadrada) $\Rightarrow A^+ = A^{-1} \Rightarrow$ $A^{-1}A = I$, $AA^{-1} = I$

Formas Cuadráticas

Sea $A \in \mathbb{R}^{n \times n}$, definimos $f(x) = x^T A x$ como la forma cuadrática asociada a la matriz A

Definimos $\bar{A} = \frac{1}{2}(A + A^T)$ como la "parte simétrica" de la matriz A

$\bar{A} = \bar{A}^T \Rightarrow \bar{A}$ es simétrica

$$f(x) = x^T A x = x^T \bar{A} x$$

Podemos acotar $f(x)$ inferiormente y superiormente por:

$$\lambda_{min}(\bar{A}) \cdot \|x\|^2 \leq x^T \bar{A} x \leq \lambda_{min}(\bar{A}) \|x\|^2$$

Positividad (de matriz cuadrada)

Sea $A \in \mathbb{R}^{n \times n}$

Si A es semidefinida positiva $\iff x^T A x \geq 0 \ \forall x \in \mathbb{R}^n$

$\iff \lambda_{min}(A) \geq 0$

Si A es definida positiva $\iff x^T A x > 0 \ \forall x \in \mathbb{R}^n$

$\iff \lambda_{min}(A) > 0$

Ganancia de una matriz (de matriz cualquiera)

Sea $A \in \mathbb{R}^{m \times n}$

Definimos la ganancia de la matriz A en la dirección x $= \dfrac{\|Ax\|}{\|x\|}$

Definimos la norma de la matriz A : $\|A\| = \max\limits_{x \neq 0} \dfrac{\|Ax\|}{\|x\|} = \sigma_{max}(A)$

$= \sqrt{\lambda_{max}(A^TA)}$

$= \sqrt{\lambda_{max}(AA^T)}$

$\min\limits_{x \neq 0} \dfrac{\|Ax\|}{\|x\|} = \sigma_{min}(A) = \sqrt{\lambda_{min}(A^TA)}$

$= \sqrt{\lambda_{min}(AA^T)}$

Número de Condición (de matriz cualquiera)

Sea $A \in \mathbb{R}^{m \times n}$

Definimos el número de condición de la matriz A:

$$K(A) = \|A\| \cdot \|A^{\dagger}\| = \frac{\sigma_{max}(A)}{\sigma_{min}(A)}$$

$K \in [1, \infty)$

Si $K(A)$ es pequeño $\Longrightarrow$ A está "bien condicionada"

Si $K(A)$ es muy grande $\Longrightarrow$ A está "mal condicionada"