Análisis Matemático para Inteligencia Artificial

Martín Errázquin (merrazquin@fi.uba.ar)

Especialización en Inteligencia Artificial

Clase 3

Colab

El mundo del Data Science y las competencias de Machine Learning cobraron mucha relevancia con la competencia de Netflix, que prometía **1M USD** a quien pudiera obtener una performance de un 10% por sobre su algoritmo de base para el problema de recomendación de películas.

En este contexto surgió una nueva variante de modelo en base a factorización de matrices.

¿Lo interesante? Con las dos clases de AM ya nos alcanza para entenderlo!

¿Por qué funciona la inversa? (1/2)

Demostremos que: $\Leftrightarrow x^T A x > 0 \quad \forall x \in \mathbb{R}^n - \{0\}$

$1{\times}n \cdot n{\times}n \cdot n{\cdot}1 \Rightarrow 1{\cdot}1$

Si $A \in \mathbb{R}^{n\times n}$ es una matriz definida positiva, entonces es invertible:

Dem.: (por método del absurdo) $\tilde{x}^T A \tilde{x} = \tilde{x}^T \cdot 0 = 0$

Supongamos que el determinante de A es cero, y llegaremos a una solución absurda. $\det(A) = 0 \Leftrightarrow \nexists A^{-1} \Leftrightarrow \exists \tilde{x} \neq 0_v \,/\, A\tilde{x} = 0$

Si $det(A) = 0$ y queremos resolver un sistema $Ax = 0$ significa que además de la solución trivial hay infinitas soluciones, es decir $rango(A) < n$, entonces

$$\exists \tilde{x} \neq 0 : A\tilde{x} = 0 \Rightarrow \tilde{x}^T A\tilde{x} = \tilde{x}^T 0 = 0$$

y entonces no se cumple que $\tilde{x}^T A\tilde{x} > 0$.

A es def. pos. $\Rightarrow \det(A) \neq 0 \Leftrightarrow \exists A^{-1}$

$\Updownarrow$

A es invertible

Demostremos que:

$A = x \cdot x^T + \lambda I_k$ **es definida positiva**, es decir, $y^T A y > 0, \; \forall y \in \mathbb{R}^k - \{0\}$

Dem.: Sea $y \in \mathbb{R}^k - \{0\}$ $\rightarrow y \neq 0$ ✓

$$y^T(x \cdot x^T + \lambda I_k)y = y^T x x^T y + \overbrace{y^T \lambda I_k y}$$

$$\lambda \cdot y^T I_k y = \lambda y^T y = \underbrace{\lambda}_{>0}\underbrace{\|y\|^2}_{>0}$$

x def. $\lambda \in \mathbb{R}^+$

(recordemos que el p.i. es $< u, v >= u^T v$), prop. asociativa,

$$< y, x >< x, y > + \lambda < y, y >= \underbrace{< x, y >^2}_{\geq 0} + \underbrace{\lambda ||y||^2}_{>0} > 0.$$

$\therefore x \cdot x^T + \lambda I_k$ es definida positiva.

$$B = \sum x_i \cdot x_i^T + \lambda I_k \Rightarrow \dots \underbrace{<x_1, y>^2}_{\geq 0} + \underbrace{<x_2, y>^2}_{\geq 0} + \dots + \underbrace{<x_n, y>^2}_{\geq 0} + \overbrace{\lambda \|y\|^2}^{>0} > 0$$

Proyección

$A^2 = A$ $\quad Av = s \quad As = s$

Sea $\mathbb{V}$ un EV y $S \subset \mathbb{V}$ un SEV. Una transformación lineal $\Pi : \mathbb{V} \to S$ es una proyección si $\Pi^2 = \Pi \circ \Pi = \Pi$ $\quad$ si $\Pi(v) = s \Rightarrow \Pi(s) = s$

Esto significa que la matriz de transformación asociada a la proyección cumple con la propiedad de idempotencia: $[\Pi]^2 = [\Pi]$.

Proyección Ortogonal

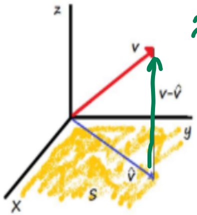
Dado $\mathbb{V}$ un EV con p.i. y $S \subset \mathbb{V}$ un SEV, el objetivo es dado $v \in \mathbb{V}$ hallar $\tilde{v} \in S$ que sea "lo más parecido posible" a v.

$v \in V$

$\tilde{v} \in S$

$\rightarrow \underbrace{v - \tilde{v}}_{v^\perp} \in S^\perp$

$$\tilde{v} \in S : \tilde{v} = arg\ min_{s \in S} ||\overbrace{v - s}^{v^\perp}||$$



Además vale que $\langle \underbrace{v - \tilde{v}}_{v^\perp}, s \rangle = 0, \ \forall s \in S$

Obs: La ortogonalidad de la proyección tiene que ver con el p.i. que se use.

Teorema de proyección

Sea $\mathbb{V}$ un EV de dimensión finita con p.i. $\langle .,. \rangle$, S un SEV. Dado $v \in \mathbb{V}$ $\underbrace{\text{existe}}_{\exists}$ un $\underbrace{\text{único}}_{!}$ $\tilde{v} \in S$ tal que

$$||v - \tilde{v}|| \leq ||v - u||, \quad \forall u \in \mathbb{S}$$

¿Cómo hallar la proyección?

Sea $\mathbb{V}$ un EV de dimensión n con p.i. $\langle .,. \rangle$, y $S \subset \mathbb{V}$ un SEV, $dim(S) = m \geq 1$, y sea $B = \{s_1, ..., s_m\}$ una BON de S. Buscamos encontrar la proyección de $\tilde{v} \in S$ de $v \in \mathbb{V}$ ($\tilde{v} = \Pi_S(v)$).

Como $\tilde{v} \in S$, $\tilde{v} = \sum_{i=1}^{m} \alpha_i s_i$ $\Rightarrow$ busco los coeficientes que minimizan $||v - \sum_{i=1}^{m} \alpha_i s_i||$. El problema puede escribirse como:

$$\Pi_S(v) = \sum_{i=1}^{m} \alpha_i s_i = B\alpha, \quad B = [s_1, ..., s_m], \ \alpha = [\alpha_1, ..., \alpha_m]^T$$

¿Cómo hallar la proyección?

Como por definición $\langle \underbrace{v - \Pi_S(v)}_{v^{\perp}}, s\rangle = 0$, $\forall s \in S$, debo resolver el siguiente sistema de ecuaciones:

para p.i. canónico $\langle x, y\rangle = y^T x = x^T y$

$$\begin{cases} \langle v - \Pi_S(v), s_1\rangle = 0 \\ \vdots \\ \langle v - \Pi_S(v), s_m\rangle = 0 \end{cases} \Rightarrow \begin{cases} s_1^T(v - B\alpha) = 0 \\ \vdots \\ s_m^T(v - B\alpha) = 0 \end{cases} \Rightarrow \overbrace{\begin{pmatrix} -s_1^T- \\ \vdots \\ -s_m^T- \end{pmatrix}}^{m\times n} \overbrace{(\underbrace{v}_{n\times 1 \checkmark} - \underbrace{B\alpha}_{\tilde{v}})}^{n\times 1} = 0$$

$$B^T(v - B\alpha) = 0 \Leftrightarrow B^T v = B^T B\alpha \Leftrightarrow \underbrace{\alpha}_{\text{coef.}} = (B^T B)^{-1} B^T v$$

vector input v

$$\Pi_S = B(B^T B)^{-1} B^T$$

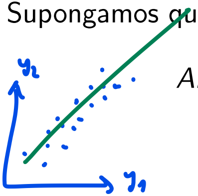
$\tilde{v} = B\cdot\alpha$

$\Rightarrow \tilde{v} = \underbrace{B\cdot(B^T B)^{-1} B^T}_{\Pi_S} v$

Observación: Si B es una BON entonces $P_{\Pi_S} = BB^T$.

$m > n$

Supongamos que tenemos un sistema sobredeterminado de la forma:



$$Ab = y, \ A \in \mathbb{R}^{m \times n}, \ b \in \mathbb{R}^n, \ y \in \mathbb{R}^m, \ m > n.$$

b es la solución de cuadrados mínimos

Como $m > n$, puede que no exista b que satisfaga todas las m ecuaciones, entonces busco la solución que más se acerque (busco $Proy_{Col(A)}y$)

$$b = \underbrace{(A^TA)^{-1}A^Ty}_{\hat{\beta}} \Rightarrow P = \underbrace{A(A^TA)^{-1}A^T}_{\Pi_s}$$

$$\hat{y} = B \cdot \alpha = A \cdot \hat{\beta}$$

Colab

Ya vimos que el ajuste por cuadrados mínimos funciona en la teoría, ahora veámoslo funcionar en la práctica!