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: $\forall x \neq \vec{0} \quad x^T A x > 0$

Si $A \in \mathbb{R}^{n \times n}$ es una matriz definida positiva, entonces es invertible:
Dem.: (por método del absurdo)

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

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 no es invertible $\Rightarrow$ A no es def pos.

$\Updownarrow$

A es def. pos $\Rightarrow$ A es invertible

$\sum x_i \cdot x_i^T + \lambda I_k = x_1 x_1^T + x_2 x_2^T + \dots + \lambda I_k$

Demostremos que:

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

$y \neq 0$

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

$1\times k \quad k\times 1 \qquad \mathbb{R} \quad k\times k$

$$y^T A y = y^T(x \cdot x^T + \lambda I_k)y = y^T x x^T y + y^T \lambda I_k y$$

$\lambda \cdot y^T \cdot I_k \cdot y$

$\lambda \cdot y^T \cdot y$

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

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

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

$\geq 0 \quad >0 \quad >0$

$\geq 0 + >0 \quad > 0$ ✓

sea $s = \Pi(v)$ $\forall s \in S$

Proyección

$\Pi(s) = s$

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

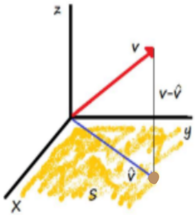
Esto significa que la matriz de transformación asociada a la proyección cumple con la propiedad de idempotencia: $[\Pi]^2 = [\Pi]$.

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.

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

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



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 $v \in \mathbb{V}$ $\underbrace{\text{existe}}_{\exists}$ $\underbrace{\text{un único}}_{!}$ $\tilde{v} \in S$ 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:

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

$$B = \begin{pmatrix} | & & | \\ s_1 & \cdots & s_m \\ | & & | \end{pmatrix} \in \mathbb{R}^{n \times m} \qquad \alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{pmatrix} \in \mathbb{R}^m$$

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

$v - \tilde{v} \in S^{\perp} \Rightarrow \forall s_i \in B \quad v - \tilde{v} \perp s_i$

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

$$\begin{cases} \langle v - \Pi_S(v), s_1 \rangle = 0 \\ \vdots \\ \langle v - \Pi_S(v), s_m \rangle = 0 \end{cases} \Rightarrow \begin{cases} s_1^T (v - B\alpha) = 0 \\ \vdots \\ s_m^T (v - B\alpha) = 0 \end{cases} \Rightarrow \overbrace{\begin{pmatrix} - s_1^T - \\ \vdots \\ - s_m^T - \end{pmatrix}}^{B^T} (v - B\alpha) = 0$$

$B = \begin{pmatrix} | & & | \\ s_1 & \cdots & s_m \\ | & & | \end{pmatrix}$

$$B^T(v - B\alpha) = 0 \Leftrightarrow B^T v = B^T B\alpha \Leftrightarrow \alpha = (B^T B)^{-1} B^T v$$

$\hat{\beta}$

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

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

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

Aplicación: Cuadrados Mínimos

$Ax = b \Rightarrow x = A^{-1} \cdot b$ sólo si $\exists A^{-1} \Leftarrow A$ cuadrada

$m > n$ $\quad$ $m < n$

Supongamos que tenemos un sistema sobredeterminado de la forma: $v \in V$

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

$\min \|\underbrace{\hat{y}}_{s \in S} - y\|$

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 $Proy_{Col(A)}y$)

$$b = \underbrace{(A^T A)^{-1} A^T}_{\hat{\beta}} y \ \Rightarrow \ P = \underbrace{A(A^T A)^{-1} A^T}_{\Pi_S}$$

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

Colab

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