Análisis Matemático para Inteligencia Artificial

Verónica Pastor (vpastor@fi.uba.ar),
Martín Errázquin (merrazquin@fi.uba.ar)

Especialización en Inteligencia Artificial

18/3/2022

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!

Demostremos que: → A simétrica $\boxed{y^t A y > 0}$ $\forall y \neq 0$

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

Dem.: (por método del absurdo) → repaso

Supongamos que el determinante de A es cero, y llegaremos a una solución absurda.

Si $det(A) = 0$ y queremos resolver un sistema $\boxed{Ax = 0}$ 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 $\tilde{x}^T A\tilde{x} > 0$.

$$u \cdot \underbrace{\left(\sum_{j \in O_i} v_j v_j^t + \lambda I_k \right)}_{A} = \sum_{j \in O} r_{ij} v_j^t$$

$$u^* = \left(\qquad \right) A^{-1}$$

¿Por qué funciona la inversa? (2/2)

Demostremos que: $k \times k$ $k \times k$

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

$\lambda > 0$ (por el modelo)

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

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

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

$<u,v> = <v,u>$

$$< y, x >< x, y > + \lambda < y, y > = < x, y >^2 + \lambda ||y||^2 > 0.$$

> 0 ≥ 0

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

Proyección Ortogonal

Proyección

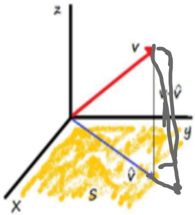
Sea $\mathbb{V}$ un EV y $S \subset \mathbb{V}$ un SEV. Una transformación lineal $\Pi : \mathbb{V} \to \mathbb{S}$ es una proyección si $\Pi^2 = \Pi \circ \Pi = \Pi$

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

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 $v \in \mathbb{V}$ existe un único $\tilde{v} \in S$ tal que

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

¿Cómo hallar la proyección?

Sea $\mathbb{V}$ un EV de dimensión 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:

$||v - \tilde{v}||$

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

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

base $\qquad \langle u, w\rangle = u^t w$ $\qquad B^T$

$$\left\{\begin{array}{c} \langle v - \Pi_S(v), s_1\rangle = 0 \\ \vdots \\ \langle v - \Pi_S(v), s_m\rangle = 0 \end{array}\right. \Rightarrow \left\{\begin{array}{c} s_1^T(v - B\alpha) = 0 \\ \vdots \\ s_m^T(v - B\alpha) = 0 \end{array}\right. \Rightarrow \overbrace{\begin{pmatrix} -s_1^T- \\ \vdots \\ -s_m^T- \end{pmatrix}}^{B^T} \cdot (v - B\alpha) = 0$$

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

? $\qquad$ desp

$$P_{\Pi_S} = B(B^T B)^{-1} B^T \qquad \mathbb{P}_{\Pi_S}(v) = B(B^T B)^{-1} B^T v$$

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

Aplicación: Cuadrados Mínimos

es decir, más ec. que incógnitas

Supongamos que tenemos un sistema sobredeterminado de la forma:

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

y 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 $Proy_{Col(A)}y$)

$$b = \underline{(A^T A)}^{-1} \underline{A^T} y \ \Rightarrow \ P = A(A^T A)^{-1} A^T$$

Colab

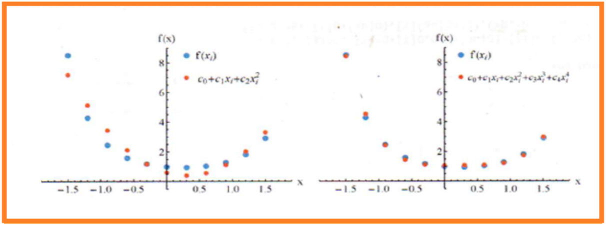
Ya vimos que el ajuste por cuadrados mínimos funciona en la teoría, ahora veámoslo funcionar en la práctica!

Ejemplo: ajustar una función por un polinomio

Sea $f : \mathbb{R} \to \mathbb{R}$ una función para un conjunto de puntos $(x_1, ..., x_n)$ a ajustar por un polinomio de grado k, $p(x) = c_0 + c_1 x + ... + c_k x^k$, $k << n$, c_i los coeficientes a ajustar.

Es decir, buscamos un polinomio que minimiza la suma de la diferencia de los cuadrados.

$$D^2 = \sum_{i=1}^{n} (f(x_i) - p(x_i))^2$$

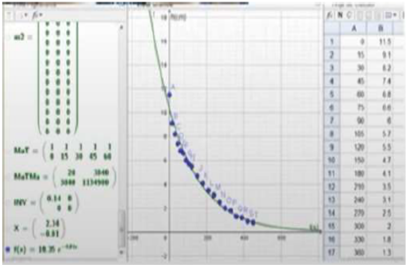


Si los datos no exhiben un comportamiento polinomial de grado k, el ajuste puede presentar oscilaciones, que requiera aumentar el grado. La matriz $A^T A$ puede estar mal condicionada al aumentar el grado.

¿Cómo varía en el tiempo la altura de la espuma de la ?

El estudio se basa en los siguientes ítems:

- Tipos de desintegración.
- Modelo funcional entre la desintegración y el tiempo.
- Relación entre el proceso de desintegración y la calidad de la cerveza.



$$\boxed{f(x) = ae^{-bx}}$$