Resumen UNI

TEMA 4

1. Aproximacion por mínimos cuadrados discreta y continua

Principio del mínimo

$A \in \mathbb{R}^{N\times N}$ es una matriz simétrica y definida positiva. $b \in \mathbb{R}^N$ y $f: \mathbb{R}^N \to \mathbb{R}^N$ es la forma cuadrática definida en cada $x \in \mathbb{R}^N$ como

$$\boxed{f(x) = \frac{1}{2}x^T A x - b^T x}$$

f alcanza su mínimo en el vector solución de $\boxed{Ax = b}$ cuyo valor mínimo es $\boxed{-1/2\, b^T A^{-1} b}$

1)* Demostración

Primero veamos que A es regular. Supongamos que $Ax=0$.
Entonces, $x^T A x = 0$ ya que cualquier vector por el vector
nulo es 0. Como A es definida positiva, $x=0$, lo cual quiere
decir que el sistema dado por $Ax=0$ es un SCD y, por
lo tanto, A es regular. Ya podemos probar que en $x = A^{-1}b$
se alcanza el mínimo:

Aplicamos la definición

Usamos que $b = Ax$

$$f(y) - f(x) = \frac{1}{2}y^T A y - b^T y - \frac{1}{2}x^T A x + b^T x =$$

Agrupamos los términos x^TAx

$$= \frac{1}{2}y^T A y - x^T A y - \frac{1}{2}x^T A x + x^T A x =$$

$$= \frac{1}{2}y^T A y + \frac{1}{2}x^T A x - x^T A y = \frac{1}{2}\left(y^T A y + x^T A x - 2x^T A y\right)$$

$$= \frac{1}{2}(y-x)^T \cdot A \cdot (y-x) \Rightarrow$$

Como A es definida positiva, $f(y) - f(x) \geq 0$, lo cual demuestra que en $x = A^{-1}b$ se alcanza el mínimo.

Aplicamos simetría de A + $(a-b)^2 = a^2 + b^2 - 2ab$

2)* Demostración

$A \in \mathbb{R}^{M\times N}$, rango$(A) = N \Leftrightarrow$ columnas de A linealmente independientes

$\Downarrow$

¿A^TA simétrica y definida positiva?

1°) Claramente A^TA es simétrica pues $(A^TA)^T = A^T(A^T)^T = A^TA$

2') Para ver que es definida positiva, sea $x \in \mathbb{R}^N$: $x \neq 0$

¿ $x^T(A^TA)x > 0$? $\Rightarrow x^T(A^TA)x = (Ax)^T Ax = \|Ax\|_2^2 \geq 0$

De hecho $\|Ax\|_2^2 = 0 \Leftrightarrow \|Ax\|_2 \Leftrightarrow Ax = 0$

$$0 = Ax = \underbrace{[a_1 | \cdots | a_N]}_{\text{columnas de } A} x = \underbrace{x_1 a_1 + \cdots + x_N a_N}_{} = 0$$

Como las columnas de A son L.I., necesariamente $x_1 = \cdots = x_N = 0$ y hemos demostrado que A^TA es definida positiva.

Aproximación por mínimos cuadrados discreta

Sea S un subespacio vectorial de $\mathbb{R}^M$ de dimensión N:

$$S = \mathcal{L}\left\{ \begin{bmatrix} a_{11} \\ \vdots \\ a_{M1} \end{bmatrix}, \ldots, \begin{bmatrix} a_{1N} \\ \vdots \\ a_{MN} \end{bmatrix} \right\} \text{ base de } S$$

A matriz con columnas los vectores de la base de S.

rango(A) = N porque son L.I. $S = \{Ax : x \in \mathbb{R}^N\}$ (en paramétricas)

Como los vectores de S son de la forma Ax, $b \in S$ si $Ax = b$ es un SCD o SCI y $b \notin S$ si $Ax = b$ es un sistema incompatible

Vamos a buscar el vector de S más próximo a uno dado $b \in \mathbb{R}^M$, es decir, hemos de hallar $x \in \mathbb{R}^N$ tal que $\forall y \in \mathbb{R}^N$ se cumpla $\|Ay - b\|_2^2 \geq \|Ax - b\|_2^2$. Tomemos $f(x) := \|Ax - b\|_2^2$ y veamos cuándo es mínima:

3)* Demostración

$$f(x) = \|Ax - b\|_2^2 = (Ax - b)^T(Ax - b) = x^T A^T A x - 2b^T Ax + b^T b =$$

$$= 2\underbrace{\left(\frac{1}{2} x^T A^T A x - (A^T b)^T x\right)}_{} + b^T b$$

Esta es justo la función que ha aparecido en el principio del mínimo, y podemos aplicar dicho principio por ser A^TA simétrica y definida positiva. f es mínima para $A^TAx = A^Tb$.

¡IMPORTANTE! $\Rightarrow$ El vector de S más cercano a b será $A((A^TA)^{-1}A^Tb)$

Otra forma de plantear el problema es el siguiente:

$$A^TAx = A^Tb$$

$$\Updownarrow$$

$$A^Tb - A^TAx = 0 \Leftrightarrow A^T(b-Ax) = 0 \Rightarrow$$ Esto equivale a imponer que el producto escalar de cada columna de A y el vector b-Ax sea 0.

$$\forall j \in \{1,\dots,N\} \Rightarrow \left\langle \begin{bmatrix} a_{1j} \\ \vdots \\ a_{Mj} \end{bmatrix}, b - Ax \right\rangle = 0$$

↳ Esto significa que b-Ax es $\perp$ a todos los vectores de la base de S ($b - Ax \in S^{\perp}$). De ahí que se llame a Ax proyección ortogonal de b sobre S.

Como $Ax = \sum_{i=1}^{N} x_i \begin{bmatrix} a_{1i} \\ \vdots \\ a_{Mi} \end{bmatrix}$, podemos reescribir el producto escalar como:

$$j \in \{1,\dots,N\} \Rightarrow \left\langle \begin{bmatrix} a_{1j} \\ \vdots \\ a_{Mj} \end{bmatrix}, b - \sum_{i=1}^{N} x_i \begin{bmatrix} a_{1i} \\ \vdots \\ a_{Mi} \end{bmatrix} \right\rangle = 0$$

Y esto se resume también en resolver un sistema.

Resumimos todo lo visto en el siguiente teorema:

Teorema de la mejor aproximación ($\mathbb{R}^M$)

Sea $A \in \mathbb{R}^{M\times N}$ con rango$(A) = N$ y sea S el subespacio:

$$S := \{Ax : x \in \mathbb{R}^N\}$$

y sea $b \in \mathbb{R}^M$. Entonces existe un único vector $x \in \mathbb{R}^N$ tal que

$$\| Ax - b\|_2 = \min\{\|Ay - b\|_2 : y \in \mathbb{R}^N\}$$

siendo Ax la mejor aproximación de b en S y proyección ortogonal de b sobre S simultáneamente, que viene caracterizado por:

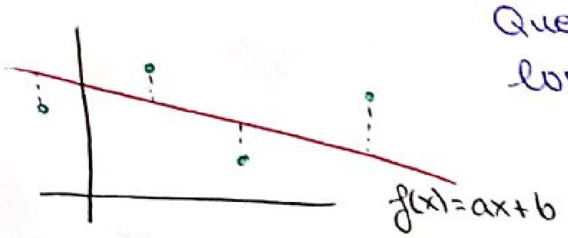
$$A^TAx = A^Tb \quad \text{(Ecuaciones normales)}$$

$$\forall j \in \{1,\dots,N\} \Rightarrow \left\langle \begin{bmatrix} a_{1j} \\ \vdots \\ a_{Mj} \end{bmatrix}, b - \sum_{i=1}^{N} x_i \begin{bmatrix} a_{1i} \\ \vdots \\ a_{Mi} \end{bmatrix} \right\rangle = 0$$

Aplicaciones

Ajuste mínimos cuadrados

Dados los datos $(x_0,y_0),(x_1,y_1),\dots,(x_M,y_M) \in \mathbb{R}^2$, buscamos una función que pase lo más cerca posible de todos los puntos a la vez.



Queremos minimizar la suma de los errores al cuadrado:

$$\sum_{i=0}^{n} (ax_i+b-y_i)^2$$

$$\Updownarrow$$

$$\left\| a\begin{bmatrix} x_0 \\ \vdots \\ x_n \end{bmatrix} + b\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} - \begin{bmatrix} y_0 \\ \vdots \\ y_n \end{bmatrix} \right\|_2^2$$

Si consideramos $S := \mathcal{L}\left\{ \begin{bmatrix} x_0 \\ \vdots \\ x_n \end{bmatrix}, \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right\}$ con $\dim(S)=2$, de la anterior expresión y comparándola con la de la mejor aproximación deducimos que el ajuste de mínimos cuadrados equivale a $P_S(y) = a\cdot x + b\cdot \mathbb{1}$

Si fuese con una parábola de la forma $f(x)=ax^2+bx+c$, mismo razonamiento: Minimizar $\sum_{i=0}^{n} (ax_i^2+bx_i+c-y_i)^2$

$$\Downarrow$$

Minimizar $\left\| a\begin{bmatrix} x_0^2 \\ \vdots \\ x_n^2 \end{bmatrix} + b\begin{bmatrix} x_0 \\ \vdots \\ x_n \end{bmatrix} + c\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} - \begin{bmatrix} y_0 \\ \vdots \\ y_n \end{bmatrix} \right\|_2^2$

Consideramos $S := \mathcal{L}\left\{ \begin{bmatrix} x_0^2 \\ \vdots \\ x_n^2 \end{bmatrix}, \begin{bmatrix} x_0 \\ \vdots \\ x_n \end{bmatrix}, \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right\}$ con $\dim(S)=3$ y la solución será $P_S(y) = a\cdot x^2 + b\cdot x + c\cdot \mathbb{1}$

Caso continuo

Definición. Dado un esp. vect. E, la aplicación $\langle\cdot,\cdot\rangle : E\times E \to \mathbb{R}$ es un <u>producto escalar</u> en E siempre que

- $\langle x,y\rangle = \langle y,x\rangle$ (Simetría)
- $\langle x+y,z\rangle = \langle x,z\rangle + \langle y,z\rangle$ (Linealidad en cada componente)
- $\langle \lambda x,y\rangle = \lambda\langle x,y\rangle$ $\quad \forall x,y,z\in E \ \ \lambda\in\mathbb{R}$
- $\langle x,x\rangle \geq 0$ con "=" $\Leftrightarrow x=0$

Y entonces E pasa a ser espacio euclídeo con producto escalar $\langle\cdot,\cdot\rangle$

Ejemplos productos escalares

- $E = \mathbb{R}^N$

$x, y \in \mathbb{R}^N \Rightarrow \langle x, y \rangle := \sum_{i=1}^{N} x_i y_i$

- $E = C([a,b])$

$f, g \in C([a,b]) \Rightarrow \langle f, g \rangle := \int_a^b f(x) g(x) dx$

Consecuencia inmediata de la definición de producto escalar:

- $\langle x, 0 \rangle = 0$
- $\left\langle \sum_{i=1}^{N} \lambda_i x_i, \sum_{j=1}^{M} \mu_j y_j \right\rangle = \sum_{i=1}^{N} \sum_{j=1}^{M} \lambda_i \mu_j \langle x_i, y_j \rangle$

Siendo los $\lambda_i, \mu_j \in E$ y $x_i, y_j \in E$

Definición de norma $\Rightarrow x \in E \Rightarrow \|x\| := \sqrt{\langle x, x \rangle}$

Ejemplos normas

- $E \in \mathbb{R}^N$

$x \in \mathbb{R}^N \Rightarrow \|x\| = \sqrt{\sum_{i=1}^{N} x_i^2}$

- $E = C([a,b])$

$f \in C([a,b]) \Rightarrow \|f\| = \sqrt{\int_a^b f(x)^2 dx}$

Aproximación en un espacio euclídeo arbitrario

Sea $(E, \langle \cdot, \cdot \rangle)$ espacio euclídeo, $v \in E$, S subespacio finito dimensional vectorial de E y $\{a_1, \dots, a_N\}$ base de S

¿Existe $u \in S$ más próximo a v?

¿Existe $u \in S : w \in S \Rightarrow \|u - v\| \leq \|w - v\|$?

$\Downarrow$ Elevamos al cuadrado

$$\|u - v\|^2 \leq \|w - v\|^2$$

$$\Updownarrow$$

$$\langle u - v, u - v \rangle \leq \langle w - v, w - v \rangle$$

$$\Updownarrow$$

$$\left.\begin{matrix} x \in \mathbb{R}^N \\ y \in \mathbb{R}^N \end{matrix}\right\} \Rightarrow \left\langle \sum_{i=1}^{N} x_i a_i - v, \sum_{j=1}^{N} x_j a_j - v \right\rangle \leq \left\langle \sum_{i=1}^{N} y_i a_i - v, \sum_{j=1}^{N} y_j a_j - v \right\rangle$$

$\Updownarrow$ ¿Existe $x \in \mathbb{R}^N$?

Desarrollamos los productos escalares

$$\Downarrow$$

$$\sum_{i=1}^{N} \sum_{j=1}^{N} x_i x_j \langle a_i, a_j \rangle + v^2 - 2 \sum_{i=1}^{N} x_i \langle a_i, v \rangle \leq \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \langle a_i, a_j \rangle + v^2 - 2 \sum_{i=1}^{N} y_i \langle a_i, v \rangle$$

$$\Downarrow$$

$$\sum_{i=1}^{N} \sum_{j=1}^{N} x_i x_j \langle a_i, a_j \rangle - 2 \sum_{i=1}^{N} x_i \langle a_i, v \rangle \leq \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \langle a_i, a_j \rangle - 2 \sum_{i=1}^{N} y_i \langle a_i, v \rangle \quad (1)$$

Consideremos ahora la matriz A y el vector b:

$$A := \begin{bmatrix} \langle a_i, a_j \rangle \end{bmatrix} \qquad b := \begin{bmatrix} \langle a_1, v \rangle \\ \vdots \\ \langle a_N, v \rangle \end{bmatrix}$$

$$x, y \in \mathbb{R}^N \Rightarrow \left\langle \sum_{i=1}^{N} x_i a_i, \sum_{j=1}^{N} y_j a_j \right\rangle \Rightarrow x^T A y$$

$$\left\langle \sum_{i=1}^{N} x_i a_i, v \right\rangle = b^T x$$

Y usando estas reformulaciones sobre (1) (mirar final de la página anterior), se nos queda la siguiente expresión:

$$¿\text{existe } x \in \mathbb{R}^N : y \in \mathbb{R}^N \Rightarrow \frac{1}{2} x^T A x - b^T x \leq \frac{1}{2} y^T A y - b^T y?$$

Volvemos a tener la función del principio del mínimo:

$$f(x) = \frac{1}{2} x^T A x - b^T x$$

Y podemos aplicar dicho principio ya que A es simétrica y definida positiva por la forma en la que la hemos definido (el producto escalar de un vector con él mismo siempre es $\langle x, x \rangle > 0$ y $= 0$ solo cuando $x = 0$). Por lo tanto, la solución sabemos que será

$Ax = b$ (ecuaciones normales).

Interpretación geométrica

$$Ax = b$$

$\Updownarrow$ (usamos que $A = [\langle a_i, a_j \rangle]$ $b = \langle a_i, v \rangle$)

$$\forall i \in \{1, \dots, N\} \quad \sum_{j=1}^{N} x_j \langle a_i, a_j \rangle = \langle a_i, v \rangle$$

$$\Updownarrow$$

$$\forall i \in \{1, \dots, N\} \Rightarrow \left\langle a_i, v - \sum_{j=1}^{N} x_j a_j \right\rangle = 0$$

$$\Updownarrow$$

$v - \sum_{j=1}^{N} x_j a_j$ es perpendicular a los vectores de la base S, por lo que $u = \sum_{j=1}^{N} x_j a_j$ es la proyección ortogonal de v sobre S.

Resumimos todo lo visto en el siguiente teorema:

Teorema de la mejor aproximación

Sean $(E, \langle \cdot, \cdot \rangle)$ un espacio euclídeo, S un subespacio vectorial finito dimensional con base $\{a_1, \ldots, a_N\}$ y sea $v \in E$. Existe un único vector $u \in S$ (mejor aproximación de v en S, proyección ortogonal de v sobre S) de forma que

$$\|u - v\| = \min\{\|w - v\| : w \in S\}$$

Las coordenadas $x \in \mathbb{R}^N$ de u en la base $\{a_1, \ldots, a_N\}$ caracterizadas por las ecuaciones normales $Ax = b$ con $A := [\langle a_i, a_j \rangle]$ y $b := [\langle a_i, v \rangle]_{i=1}^{N} \in \mathbb{R}^N$ o bien por $\forall i \in \{1, \ldots, N\} \Rightarrow \langle a_i, v - \sum_{j=1}^{N} x_j a_j \rangle = 0$

Aproximación uniforme

Teorema de Weierstrass: $\mathbb{P}$ denso en $C[a,b]$ (norma del máximo)

$$\Downarrow$$

$$\forall \varepsilon > 0, f \in C[a,b] \Rightarrow \exists p \in \mathbb{R} : \|f - p\|_\infty < \varepsilon$$

Polinomios de Berstein: $N \geq 1$, $f \in C([0,1])$, el polinomio de Berstein $B_N f$ de orden N para f está definido en cada $x \in [0,1]$ como

$$B_N f(x) := \sum_{i=0}^{N} f(i/N) \binom{N}{i} x^i (1-x)^{N-i}$$

Versión numérica del Teorema de Weierstrass

$f \in C([0,1])$

$$\lim_{N \to \infty} \|f - B_N f\| = 0$$

* (Para el ejercicio 4 de la relación) En concreto, si $f \in C^1([0,1])$, $\|f - B_N f\|_\infty \leq \frac{\|f'\|_\infty}{\sqrt{N}}$