Cuadrados mínimos

Juan Pablo Pinasco (`jpinasco@dm.uba.ar//gmail.com`)

Departamento de Matemática e IMAS,
FCEyN, UBA - CONICET

2020

Hoy:

- Cuadrados mínimos

Próxima:

- Ec no lineales: 65-
- bisec:
- Newton Raph: -71

Parte I

(a) Cuadrados mínimos

1.- Problema (a)

Dada una tabla de datos,

x	x_1	x_2	$\cdots$	x_n
y	y_1	y_2	$\cdots$	y_n

sabemos *interpolarlos*, encontrar una función p (en general, un polinomio) tal que $p(x_i) = y_i$.

Hay varios casos donde esto falla:

- valores de x_i muy juntos, que provoca oscilaciones muy grandes.
- valores repetidos de x: típico en experimentos donde se fija un x y se mide el y,
- n demasiado grande implica polinomios inmanejables que oscilan a lo bestia.

Cuadrados mínimos

Buscamos una función (por ejemplo, un polinomio de grado $m \ll n$) que minimice

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

Generalización: dadas $\phi_0, \ldots, \phi_m$, funciones conocidas, hallar los parámetros $\{a_i\}_{1=0:m}$ tal que se minimice

$$\sum_{i=1}^{n} (a_0\phi_0(x_i) + a_1\phi_1(x_i) + \cdots + a_m\phi(x_i) - y_i)^2$$

También podemos cambiar la norma 2 por otra, minimizar

$$\sum_{i=1}^{n} |a_0\phi_0(x) + a_1\phi_1(x_i) + \cdots + a_m\phi(x_i) - y_i|,$$

$$\max_i |a_0\phi_0(x) + a_1\phi_1(x_i) + \cdots + a_m\phi(x_i) - y_i|$$

que serían la norma 1 y la ∞ entre el vector de los datos y_i y el vector obtenido al evaluar la función en x_i.

Supongamos que $\phi_i(x) = x^i$, estamos buscando un polinomio de grado m que aproxime los datos. Si $p(x) = a_m x^m + \cdots + a_1 x + a_0$, tenemos

$$Ax = \begin{pmatrix} 1 & x_1 & \cdots & x_1^m \\ 1 & x_2 & \cdots & x_2^m \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^m \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = b$$

Este sistema $Ax = b$ no puede resolverse, la matriz no es inversible (ni siquiera es cuadrada), pero nos permite plantear el problema en una forma que la proyección ortogonal lo resuelve:

Sea $S = \{y \in \mathbb{R}^n \ : \ y = Ax \text{ para } x \in \mathbb{R}^m\}$

Entonces, queremos $y \in S$ tal que

$$\|y - b\| \leq \|s - b\| \text{ para todo } s \in S.$$

Para resolverlo necesitamos un lema:

Lema: Sea $A \in \mathbb{R}^{n\times m}$, $y \in \mathbb{R}^n$, $x \in \mathbb{R}^m$, $\langle \cdot, \cdot \rangle$ el producto interno usual de $\mathbb{R}^n$ o $\mathbb{R}^m$. Entonces

$$\langle A^t y, x\rangle = \langle y, Ax\rangle$$

Dem: cuenta,

$$\begin{aligned}\langle A^t y, x\rangle =& \sum_{i=1}^{m}\Big(\sum_{j=1}^{n} a_{ji}y_j\Big)x_i\\ =& \sum_{j=1}^{n} y_j\Big(\sum_{i=1}^{m} x_i a_{ji}\Big)\\ =& \langle y, Ax\rangle\end{aligned}$$

Teorema

Sea $A \in \mathbb{R}^{n\times m}$, $b \in \mathbb{R}^n$, $x_0 \in \mathbb{R}^m$. Entonces, son equivalentes:

1. *x_0 minimiza $\|Ax - b\|$.*
2. *x_0 es solución de $A^tAx = A^tb$.*

Si las columnas de A son l.i., la solución x_0 de $A^tAx = A^tb$ existe y es única.

5.- Demostración (equivalencia)

Sabemos de la clase anterior que minimizar $\|y-b\|$ en S es encontrar un y tal que

$$\langle b-y, s\rangle = 0 \quad \text{para todo } s \in S.$$

Como los elementos de S son de la forma $s = Ax$, existe un $x_0 \in S$ que minimiza si y solo si

$$\langle b - Ax_0, Ax\rangle = 0 \quad \text{para todo } x \in \mathbb{R}^m$$

Por el Lema,

$$\langle A^t(b - Ax_0), x\rangle = 0 \quad \text{para todo } x \in \mathbb{R}^m$$

pero entonces tiene que ser $A^t(b-Ax_0) = 0$, es decir

$$A^tAx_0 = A^tb.$$

Veamos la unicidad. Sea $x = (x_1, \dots, x_m)$.

El vector $Ax = \sum_{j=1}^{m} x_j A_j$, donde A_j es la columna j de A.

Si las columnas son l.i., la única forma de que Ax sea cero es que $x = 0$.

Pero esto implica que $A^t A$ es inversible, ya que si $A^t Ax = 0$, tenemos

$$0 = \langle A^t Ax, x\rangle = \langle Ax, Ax\rangle = \|Ax\|^2,$$

así que $Ax = 0$, y como las columnas era l.i., $x = 0$.
Entonces, el sistema homogéneo $A^t Ax = 0$ tiene una solución $x = 0$, y por lo tanto la matriz es inversible.

Si $\langle x, y\rangle = \sum_{j=1}^{n} w_j x_j y_j$, tenemos que x_0 minimiza $\|Ax - b\|_w$ si y solo si es solución de $A^t D_w Ax = A^t D_w b$, pensamos los w_j como una matriz diagonal multiplicando $Ax - b$.

Parte II

(b) Aproximación polinomial

Recordemos que dado S, un subespacio de dimensión finita de V, podemos definir $P : V \to S$, donde Px es el único elemento de S que cumple

$$\langle x - Px, s\rangle = 0 \qquad \forall s \in S,$$

y Px se llama la proyección ortogonal de x en S.

Sea $V = C[a,b]$, $S = P_n = \langle a, x, x^2, \dots, x^n\rangle$, y

$$\langle f, g\rangle = \int f(x)g(x)w(x)dx$$

Construimos una base ortonormal vía G-S,

$$q_0(x) = 1, \qquad p_0 = q_0(x)/\|q_0\|$$

$$q_k(x) = x^k - \sum_{i=1}^{k-1} \langle x^k, p_i\rangle p_i(x), \qquad p_k = q_k(x)/\|q_k\|$$

Teorema

Los $\{q_k\}$ son una base ortogonal, son mónicos, y $\langle q_i, p\rangle = 0$ si $gr(p) < i$.

Teorema

Si $f \in C[a,b]$*, el polinomio* $p_n^* \in P_n$ *que satisface*

$$\|f - p_n^*\| \leq \|f - p\| \qquad \forall p \in P_n$$

es la proyección ortogonal

$$p_n^* = \sum_{i=0}^{n} \langle f, p_i \rangle p_i$$

No hay que hacer nada, sale de los teoremas de antes.

Teorema

Si $f \in C[a,b]$, *y* $\int_a^b w dx < C$, *entonces* $p_n^* \to f$ *cuando* $n \to \infty$ *en* $\| \cdot \|_2$.

Dem.: por Weierstrass, existe una sucesión de polinomios p_n tales que

$$\|f - p_n\|_\infty \leq \frac{1}{n}$$

Por el teorema anterior, existe p_n^* de grado menor o igual que p_n tal que

$$\begin{aligned}
\|f - p_n^*\|_2^2 \leq & \|f - p_n\|_2^2 \\
& = \int_a^b w(x)|f(x) - p_n(x)|^2 dx \\
& \leq \|f - p_n\|_\infty^2 \int_a^b w(x) dx \\
& \leq \frac{1}{n^2} \int_a^b w(x) dx \\
& \to_{n \to \infty} 0
\end{aligned}$$

$\square$

Teorema (Parseval)

Si $f \in C[a,b]$, $\|f\|^2 = \lim_{n\to\infty} \sum_{j=1}^{n} \langle f, p_j\rangle^2$.

Dem: como $p_n^* = \sum_{i=0}^{n} \langle f, p_i\rangle p_i$, y como son ortonormales,

$$\|p_n^*\|^2 = \sum_{i=0}^{n} \langle f, p_i\rangle^2.$$

Por otra parte, $f - p_n^*$ y p_n^* son ortogonales, así que

$$\|f\|^2 = \|f - p_n^*\|^2 + \|p_n^*\|^2 = \|f - p_n^*\|^2 + \sum_{i=0}^{n} \langle f, p_i\rangle^2.$$

Ahora, cuando $n \to \infty$, tenemos

$$\lim_{n\to\infty} \|f\|^2 = \lim_{n\to\infty} \left(\|f - p_n^*\|^2 + \sum_{i=0}^{n} \langle f, p_i\rangle^2 \right) = 0 + \lim_{n\to\infty} \sum_{i=0}^{n} \langle f, p_i\rangle^2.$$

Teorema

Si $\langle xf, g\rangle = \langle f, xg\rangle$*, los polinomios ortogonales mónicos* q_n *satisfacen*

$$q_n(x) = (x - a_n)q_{n-1}(x) - b_n q_{n-2}(x) \qquad n \geq 2,$$

$$a_n = \frac{\langle xq_{n-1}, q_{n-1}\rangle}{\langle q_{n-1}, q_{n-1}\rangle}, \qquad b_n = \frac{\langle q_{n-1}, q_{n-1}\rangle}{\langle q_{n-2}, q_{n-2}\rangle}.$$

Dem: Sea $n \geq 2$. Como 0 es raíz de $q_n(x) - q_n(0)$, podemos escribir

$$q_n(x) - q_n(0) = xr_{n-1}$$

con r_{n-1} mónico, de grado menor o igual a $n-1$ (pues q_n es mónico).

Luego,

$$\begin{aligned} q_n(x) =& xr_{n-1}(x) + q_n(0) \\ =& xq_{n-1}(x) + x[r_{n-1}(x) - q_{n-1}(x)] + q_n(0) \\ =& xq_{n-1}(x) + \sum_{j=0}^{n-1} \beta_j q_j(x) \end{aligned}$$

porque $r_{n-1}(x) - q_{n-1}(x)$ es de grado menor o igual a $n-2$, ya que son mónicos.

Teníamos $q_n(x) = xq_{n-1} + \sum_{j=0}^{n-1} \beta_j q_j$.

Separemos el q_{n-1} de la sumatoria, y calculemos $\langle q_n, q_i \rangle$ para $i < n-2$:

$$0 = \langle q_n, q_i \rangle = \langle (x + \beta_{n-1}) q_{n-1} + \sum_{j=0}^{n-2} \beta_j q_j, q_i \rangle$$
$$= \langle x q_{n-1}, q_i \rangle + \beta_i \langle q_i, q_i \rangle$$

Como $\langle xf, g \rangle = \langle f, xg \rangle$, y xq_i es de grado menor o igual a $n-2$,

$$\langle xq_{n-1}, q_i \rangle = \langle q_{n-1}, xq_i \rangle = 0.$$

Entonces, $\beta_i = 0$ si $i \leq n-2$, y

$$q_n(x) = (x - a_n) q_{n-1}(x) - b_n q_{n-2}(x) \qquad n \geq 2,$$

Los coeficientes se obtienen calculando el producto interno contra q_{n-1} y q_{n-2}. $\square$