Apt 1.3, Pràctica 3 ALN: Mínims quadrats

Camelia Brumar
Leonor Lamsdorff-Galagane

08 May 2017

1 Demostració que $p_M(x)$ és el polinomi "òptim"

Sigui E un espai euclidià, i $F \subseteq E$ un sub-espai vectorial de dimensió finita. Procedirem a provar que, $\forall x \in E$ es té:

$$\|x - \pi_F(x)\| < \|x - u\|, \quad \forall u \in F, \quad u \neq \pi_F(x)$$

És a dir, que la projecció ortogonal $\pi_F(x)$ és la millor aproximació de x en F.

$$x - u = (x - \pi_F(x)) + (\pi_F(x) - u)$$

Com els dos sumands de la dreta són ortogonals (el primer pertany a $F^\perp$ i el segon a F), s'aplica el Teorema de Pitàgores:

$$\|x - u\|^2 = \|x - \pi_F(x)\|^2 + \|\pi_F(x) - u\|^2$$

Posat que $u \neq \pi_F(x)$ i $\|\pi_F(x) - u\|^2 > 0$, es té:

$$\|x - u\|^2 > \|x - \pi_F(x)\|^2$$

Com la norma és un nombre real no negatiu, ja hem acabat. Una vegada provat aquest resultat, cal veure quina forma prenen els vectors $x \in M_{n\times 1}(\mathbb{R})$ que minimitzen la norma $\|Ax - b\|$, on $A \in M_{m\times n}(\mathbb{R})$ i $b \in M_{m\times 1}(\mathbb{R})$, i veurem que són les solucions del sistema

$$A^\top A x = A^\top b$$

Si prenem $F = [u_1, ..., u_n]$ com el sev generat per les columnes de A, observem que Ax són combinacions lineals de les columnes de A, és a dir, $\{Ax\}_{x\in M_{n\times 1}(\mathbb{R})} \in F$.

x^* minimitza la norma $\|Ax - b\| \iff Ax^* = \pi_F(b)$, pel resultat anterior. Però a la vegada, $Ax^* = \pi_F(b) \iff < u_i, Ax^* > = < u_i, b > \forall i = 1 \div n$ (caracterització coneguda de la projecció ortogonal). Per últim,

$$< u_i, Ax^* > = < u_i, b > \forall i = 1 \div n \iff A^\top A x^* = A^\top b$$

Conclusió: les solucions del sistema $A^\top Ax = A^\top b$ minimitzen la norma $\|Ax - b\|$. $\square$

Aleshores, si volem trobar un polinomi de grau m que aproximi uns vectors de punts $x = (x_0, ..., x_n)$ i $y = (y_0..., y_n)$, estem buscant un polinomi $a_0x^m + a_1x^{m-1} + ... + a_{m-1}x^1 + a_m$ que satisfaci:

$$a_0x_0^m + a_1x_0^{m-1} + ... + a_{m-1}x_0^1 + a_m = y_0$$

$$a_0x_1^m + a_1x_1^{m-1} + ... + a_{m-1}x_1^1 + a_m = y_1$$

$$...$$

$$a_0x_n^m + a_1x_n^{m-1} + ... + a_{m-1}x_n^1 + a_m = y_n$$

Matricialment, equival a resoldre el sistema

$$\begin{pmatrix} x_0^m & x_0^{m-1} & \dots & x_0^1 & 1 \\ x_1^m & x_1^{m-1} & \dots & x_1^1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n^m & x_n^{m-1} & \dots & x_n^1 & 1 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{pmatrix}$$

Com, en general, $m < n$, no hi ha solució, però gràcies als apartats anteriors sabem que el polinomi $a_0^*x^m + a_1^*x^{m-1} + ... + a_{m-1}^*x^1 + a_m^*$ obtingut com a solució de les equacions normals és el que minimitza la norma $\|Ax - b\|$.

2 Justificació del mètode

A continuació justificarem el mètode utilitzat per trobar les solucions de les equacions normals i l'explicarem breument.

En aquesta pràctica hem utilitzat els mètodes d'ortogonalització, que aconsegueixen la reducció a forma triangular mitjançant l'ús de matrius ortogonals. Aquests mètodes estan basats en la descomposició QR de matrius. Hem triat aquests mètodes perquè respecte d'altres, com Cholesky (si la matriu A és simètrica) o LU, presenta certs avatatges. Per una banda, en fer Cholesky, s'haurien de resoldre dos sistemes lineals (a més d'altres càlculs anteriors), mentre que en fer QR només hem de resoldre un sistema triangular superior. Per altra banda, en comparar la QR amb la LU, la primera és molt més estable en quant als errors respecte la segona, tot i que és més ineficient pel que fa el temps computacional.

Com en els mètodes gaussians, s'observen dues parts en el procés d'ortogonalització:

- En la primera, es factoritza A en la forma:

$$A = QR,$$

  on Q és ortogonal i R triangular superior.

- En la segona, s'acaba la resolució del sistema $Ax = QRx = b$, resolent el sistema triangular superior $Rx = Q^\top b$ pel mètode de substitució cap enrere. Arribem a aquest sistema senzill seguint els següents passos:

$$A^\top Ax = A^\top b\ ,$$

  com que $A = QR$

$$(QR)^\top QRx = (QR)^\top b$$
$$R^\top Q^\top QRx = R^\top Q^\top b$$

  sabent que Q és una matriu ortogonal, i.e, $Q^\top = Q^{-1}$, obtenim

$$R^\top Rx = R^\top Q^\top b$$

  i finalment multipliquem per la inversa de $R^\top$ a les dues bandes de la igualtat, aconseguint el sistema

$$Rx = Q^\top b\ .$$

Aquesta factorització l'hem realitzat amb el mètode d'ortogonalització modificat de Gram-Schmidt.

Mètode d'ortogonalització modificat de Gram-Schmidt

Començant amb $A_1 = A$, una vegada coneguda

$$A_k = \left(q_1\ \dots\ q_{k-1}\ a_k^{(k)}\ \dots\ a_n^{(k)} \right)$$

amb columnes $q_j\ \ (j = 1 \div k-1)$ i $a_s^{(k)}\ \ (s = k \div n)$ complint amb les relacions d'ortogonalitat

$$q_j^\top q_l = \delta_{jl},\ q_j^\top a_s = 0\ (j, l = 1 \div k-1,\ s = k \div n)\,,$$

es normalitza la k-èssima columna i s'ortogonalitzen, respecte a ella, totes les que la segueixen:

$$r_{kk} = \left|\left|\ a_k^{(k)}\ \right|\right|_2,\ q_k = \frac{a_k^{(k)}}{r_{kk}}\ ;$$

$$r_{ks} = q_k^\top a_s^{(k)}\ ,\ \ a_s^{(k+1)} = a_s^{(k)} - r_{ks} q_k\ \ \ (s = k+1 \div n)\,.$$

S'obté

$$A_{k+1} = \left(q_1\ \dots\ q_k\ a_{k+1}^{(k+1)}\ \dots\ a_n^{(k+1)} \right)\ ,$$

verificant les mateixes condicions d'ortogonalitat anteriors, substituint k per $k+1\ \ (k = 1 \div n)$.
Després de n passos,

$$A_{n+1} = (q_1\ q_2 \dots q_n)$$

és una matriu ortogonal perquè $q_j^T q_l = \delta_{jl}\ \ (j, l = 1 \div n)$.

Les matrius $Q = A_{n+1}$ i $R = r_{(ks)}$ formen una factorització QR de A.