SME0305 - 2016

Gustavo C. Buscaglia / Roberto F. Ausas

ICMC - Ramal 738176, gustavo.buscaglia@gmail.com
ICMC - Ramal 736628, rfausas@gmail.com

Sistemas lineares sobredeterminados

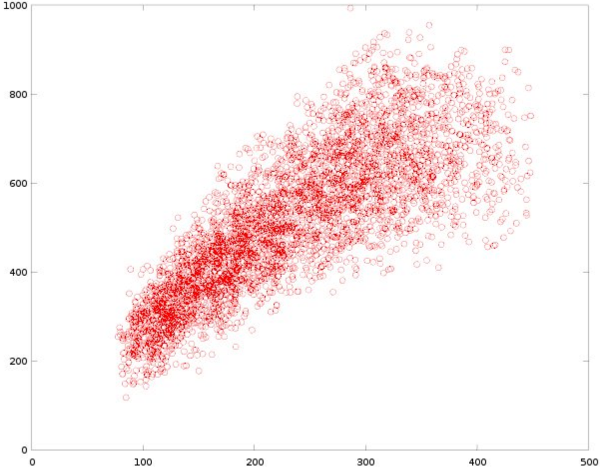
Matrizes não quadradas

- Onde aparecem, no dia a dia de um engenheiro, matrizes não quadradas?

#	Terr m^2	Constr m^2	Ano	suites	quartos	banh	Plantas	Piscina m	Vagas	Seg24h 1=sim	Preço kR$
1	202	140	1998	0	3	3	2	0	2	0	423
2	250	137	2011	1	2	4	1	0	3	1	611
3	156	102	2001	1	1	3	1	0	1	1	354
4	353	182	2004	3	0	4	1	12	3	0	712
5	198	145	1983	2	1	5	1	0	2	0	387
6	376	251	2007	3	1	5	2	14	3	1	971
7	242	165	2015	1	3	4	2	12	3	1	765
8	177	133	1976	0	3	3	1	0	1	1	313
9	298	223	1997	3	0	5	1	0	2	1	789
10	422	351	2004	3	2	7	2	18	3	1	1310
11	202	140	1998	0	3	3	2	0	2	0	423
12	250	137	2011	1	2	4	1	0	3	1	611
13	156	102	2001	1	1	3	1	0	1	1	354
14	353	182	2004	3	0	4	1	12	3	0	712
15	198	145	1983	2	1	5	1	0	2	0	387
16	376	251	2007	3	1	5	2	14	3	1	971
17	242	165	2015	1	3	4	2	12	3	1	765
18	177	133	1976	0	3	3	1	0	1	1	313
19	298	223	1997	3	0	5	1	0	2	1	789
20	422	351	2004	3	2	7	2	18	3	1	1310
21	202	140	1998	0	3	3	2	0	2	0	423
22	250	137	2011	1	2	4	1	0	3	1	611
23	156	102	2001	1	1	3	1	0	1	1	354
24	353	182	2004	3	0	4	1	12	3	0	712
25	198	145	1983	2	1	5	1	0	2	0	387
26	376	251	2007	3	1	5	2	14	3	1	971
27	242	165	2015	1	3	4	2	12	3	1	765
28	177	133	1976	0	3	3	1	0	1	1	313
29	298	223	1997	3	0	5	1	0	2	1	789
30	422	351	2004	3	2	7	2	18	3	1	1310
31	202	140	1998	0	3	3	2	0	2	0	423
32	250	137	2011	1	2	4	1	0	3	1	611
33	156	102	2001	1	1	3	1	0	1	1	354
34	353	182	2004	3	0	4	1	12	3	0	712
35	198	145	1983	2	1	5	1	0	2	0	387

- Sistemas sobredeterminados aparecem quando se deseja fazer sentido desses dados.
- Poderia se procurar uma relação entre superfície construída e preço, por exemplo. Graficando as colunas uma como função da outra,

```
plot(Dados(:,3),Dados(:,12))
```



Preço vs. Superfície construída

- O gráfico mostra uma certa tendência, mas não fornece uma fórmula para realizar uma estimativa rápida.

  É popular procurar fórmulas lineares (afins), do tipo

$$\text{preço} \simeq \text{constante}_1 + \text{constante}_2 \times \text{supconstr}$$

  Isto é, procurar k_1 e k_2 tais que

$$\begin{array}{rcl} k_1 + S_1\, k_2 & = & P_1 \\ k_1 + S_2\, k_2 & = & P_2 \\ \cdots & & \cdots \\ k_1 + S_m\, k_2 & = & P_m \end{array} \Leftrightarrow \begin{pmatrix} 1 & S_1 \\ 1 & S_2 \\ \cdots & \cdots \\ 1 & S_m \end{pmatrix} \begin{pmatrix} k_1 \\ k_2 \end{pmatrix} = \begin{pmatrix} P_1 \\ P_2 \\ \cdots \\ P_m \end{pmatrix}$$

  Eis aqui um **sistema sobredeterminado**.

- Está claro (do gráfico) que k_1 e k_2 tais que o sistema se cumpra exatamente **não existem**. Caso contrário todos os pontos estariam sobre uma mesma reta.

  Qual seria a resposta “correta”? Em que sentido?

  Como podemos definir a “menos pior” das retas possíveis?

- O gráfico mostra outras duas coisas:
  - A tendência dos dados não é exatamente uma reta, se observa uma concavidade para abaixo...
  - Para cada valor de `supconstr` existe uma grande dispersão...

- Para modelar a concavidade da tendência, é popular ajustar uma quadrática,

$$\text{preço} \simeq k_1 + k_2 \times \text{supconstr} + k_3 \times \text{supconstr}^2$$

$$\begin{array}{c} k_1 + S_1 k_2 + S_1^2 k_3 = P_1 \\ k_1 + S_2 k_2 + S_2^2 k_3 = P_2 \\ \ldots \\ k_1 + S_m k_2 + S_m^2 k_3 = P_m \end{array} \Leftrightarrow \begin{pmatrix} 1 & S_1 & S_1^2 \\ 1 & S_2 & S_2^2 \\ \ldots & \ldots & \ldots \\ 1 & S_m & S_m^2 \end{pmatrix} \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix} = \begin{pmatrix} P_1 \\ P_2 \\ \ldots \\ P_m \end{pmatrix}$$

Eis aqui mais um **sistema sobredeterminado**.

Ele também não tem chance de ter solução.

- A grande dispersão pode vir de considerarmos apenas a área construida, havendo outros fatores importantes. Podemos suspeitar que é necessário incluir na estimativa a **superfície do terreno**:

$$\text{preço} \simeq k_1 + k_2 \times \text{supconstr} + k_3 \times \text{supterr}$$

$$\left(\begin{array}{ccc} 1 & S_1 & T_1 \\ 1 & S_2 & T_2 \\ ... & ... & ... \\ 1 & S_m & T_m \end{array} \right) \left(\begin{array}{c} k_1 \\ k_2 \\ k_3 \end{array} \right) = \left(\begin{array}{c} P_1 \\ P_2 \\ ... \\ P_m \end{array} \right)$$

Eis aqui mais um **sistema sobredeterminado**.

Álgebra linear relevante

É conveniente revisar alguns conceitos básicos:
Seja $A \in \mathbb{R}^{m \times n}$, com colunas c_j e linhas ℓ_i^T:

$$A = (a_{ij}) = (c_1 | c_2 | \dots | c_n) = \begin{pmatrix} \ell_1^T \\ \dots \\ \ell_m^T \end{pmatrix}$$

A imagem de A é o espaço gerado pelas colunas de A,

$$\mathrm{Im}(A) = \mathrm{col}(A) = \{\text{comb. lineares de colunas de } A\}$$

Verifiquemos:

$$
\begin{aligned}
A\,v &= (c_1|\ldots|c_n)\begin{pmatrix} v_1 \\ v_2 \\ \ldots \\ v_n \end{pmatrix} \\
&= (c_1|\ldots|c_n)\begin{pmatrix} v_1 \\ 0 \\ \ldots \\ 0 \end{pmatrix} + \ldots + (c_1|\ldots|c_n)\begin{pmatrix} 0 \\ 0 \\ \ldots \\ v_n \end{pmatrix} \\
&= v_1\,c_1 + v_2\,c_2 + \ldots + v_n\,c_n \quad \square
\end{aligned}
$$

A matriz A transforma o espaço de linhas $\text{lin}(A) \subset \mathbb{R}^n$ no espaço de colunas $\text{col}(A) \subset \mathbb{R}^m$.

Isto é bem evidente nas matrizes simples geradas por dois vetores $a \in \mathbb{R}^m$ e $z \in \mathbb{R}^n$:

$$A = a\, z^T = \begin{pmatrix} a_1 \\ \dots \\ a_m \end{pmatrix} (z_1 \dots z_n) = \begin{pmatrix} a_1 z_1 & \dots & a_1 z_n \\ \dots & \dots & \dots \\ a_m z_1 & \dots & a_m z_n \end{pmatrix} = (a_i\, z_j)$$

$$\text{col}(A) = \{\alpha\, a\,,\ \alpha \in \mathbb{R}\}, \qquad \text{lin}(A) = \{\alpha\, z\,,\ \alpha \in \mathbb{R}\}\ .$$

Se $v \perp z$ então $Av = 0$. $\square$
Se $v = \alpha z$ então $Av = \alpha \|z\|_2^2\, a$. $\square$

No caso do sistema sobredeterminado $Ax = b$, existe solução se e só se $b \in \text{col}(A) = \text{Im}(A)$.

$$b \notin \text{col}(A) \quad \Rightarrow \quad r(x) = Ax - b \neq 0 \ , \ \forall\, x \in \mathbb{R}^n$$

Mas isto não quer dizer que todos os x estejam **igualmente** errados!

Como escolher um deles?

A distância entre dois vetores v e w, em um espaço vetorial com produto escalar

$$v \cdot w = v^T\, G\, w$$

é dada por

$$\mathrm{dist}(v, w) = \|v - w\|_G = \sqrt{(v - w)^T G (v - w)}\ .$$

Todo produto escalar corresponde a uma certa matriz G, que deve ser simétrica e positiva definida ($v^T G v > 0\ \forall\, v \neq 0$).

O produto usual (norma $\|\cdot\|_2$) corresponde a $G = I$.

Equações normais

Resulta natural escolher a **solução de quadrados mínimos**:

A solução x^* do sistema sobredeterminado $Ax = b$ é aquela que minimiza $\|Ax - b\|_G$ para o produto escalar G de $\mathbb{R}^m$, i.e.,

$$\|Ax^* - b\|_G \leq \|Ax - b\|_G \qquad \forall\, x \in \mathbb{R}^m \,.$$

Quando poderia aparecer um caso de $G \neq I$? Um caso é quando se da um peso diferente w_i a cada equação (confiabilidade da medição?).

$$G = \mathrm{diag}(w_1^2, w_2^2, \ldots, w_m^2)$$

Teorema (Equações normais):

$$A^T G A x^* = A^T G\, b\ .$$

Prova:

$$\begin{aligned}
\|Ax - b\|^2 &= (Ax-b)^T G (Ax-b) \\
&= (Ax^* + Ad - b)^T G (Ax^* + Ad - b) \\
&= \|Ax^* - b\|_G^2 + 2d^T (A^T G A x^* - A^T G b) + \|Ad\|_G^2 \quad \square
\end{aligned}$$

Obs: Quando A é de posto completo $n < m$ a matriz $M = A^T G A$ (simétrica) é definida positiva. **Pode ser resolvida por fatoração de Cholesky, mas frequentemente está mal condicionada e é instável ao arredondamento.**

Exemplo[1] *:* Aritmética de três dígitos. Produto usual, $G = I$.

$$A = \begin{pmatrix} 1.07 & 1.10 \\ 1.07 & 1.11 \\ 1.07 & 1.15 \end{pmatrix} \longrightarrow A^T A = \begin{pmatrix} 3.43 & 3.60 \\ 3.60 & 3.76 \end{pmatrix}$$

que não é definida positiva: $(-1\ 1)\ A^T A\ (-1\ 1)^T = -0.01$. Nesse caso o algoritmo de Cholesky falha.

[1] Tomado de Numerical Linear Algebra, de R. Rannacher, Lecture Notes WS 2013/2014, Heidelberg University.

Fatoração QR completa

Teorema: Toda matriz $\mathbf{A} \in \mathbb{R}^{m \times n}$ possui uma fatoração

$$A = Q\,R \quad \longleftrightarrow \quad \texttt{[Q R] = qr(A)}$$

onde $Q \in \mathbb{R}^{m \times m}$ é ortogonal e $R \in \mathbb{R}^{m \times n}$ é triangular (trapezoidal) superior, com $r_{ii} \geq 0 \; \forall i$.

Se o posto de A é n, então a decomposição é única, e $r_{ii} > 0 \; \forall i$

As colunas (e linhas) de Q poderiam ser calculadas ortogonais para qualquer produto escalar G. Vamos assumir que $G = I$, que é o que Octave considera.

A maneira mais clássica de obter Q e R é o algoritmo de Gram-Schmidt (infelizmente instável):

```
function [Q,R] = clgs(A)
[m,n] = size(A);
Q = zeros(m,n);
R = zeros(n,n);
for j=1:n
  V = A(:,j);
  for i=1:j-1
    R(i,j) = Q(:,i)'*A(:,j);
    V = V - R(i,j)*Q(:,i);
  end
  R(j,j) = norm(V);
  Q(:,j) = V/R(j,j);
end
```

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{pmatrix} = \begin{pmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & q_{34} \\ q_{41} & q_{42} & q_{43} & q_{44} \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \\ 0 & 0 & r_{33} \\ 0 & 0 & 0 \end{pmatrix}$$

$$\begin{aligned} c_1(A) &= r_{11}\, c_1(Q) \\ c_2(A) &= r_{12}\, c_1(Q) + r_{22}\, c_2(Q) \\ c_3(A) &= r_{13}\, c_1(Q) + r_{23}\, c_2(Q) + r_{33}\, c_3(Q) \end{aligned}$$

Primeira coluna de Q: Base de span$(c_1(A))$.
Segunda coluna de Q: Base de span$(c_1(A), c_2(A))$.
...

Notar que post-multiplicar Q por R corresponde a fazer operações com as colunas.

- As últimas $m - n$ colunas de Q são irrelevantes no produto, já que se multiplicam pelas últimas $m - n$ linhas de R, que são zero.
- As n primeiras colunas de Q (se $\text{rank}(A) = n$) são base ortonormal de $\text{col}(A)$.
- As $n - m$ últimas colunas de Q (se $\text{rank}(A) = n$) são base ortonormal de $\text{col}(A)^{\perp}$.
- Se as n colunas de A não são linearmente independentes, então algum dos r_{ii} é zero.

Utilidade: Dados n vetores de dados, podemos saber se são linearmente independentes e, se sim, construir o projetor $\Pi = Q_1 Q_1^T$, onde $Q_1 \in \mathbb{R}^{m\times n}$ consiste das primeiras n colunas de Q.

A matriz Π é tal que Πv é a projeção ortogonal de v sobre $\mathrm{col}(A)$ (o vetor de $\mathrm{col}(A)$ mais próximo de v, em norma $\|\cdot\|_2$).

Exemplo: No caso em que A tem uma coluna só, c_1, temos

$$Q_1 = \frac{1}{\|c_1\|_2} c_1 \ , \qquad \Pi = \frac{1}{\|c_1\|_2^2} c_1 \, c_1^T \ , \qquad \Pi v = \frac{c_1^T v}{\|c_1\|_2^2} c_1$$

que é claramente a projeção ortogonal sobre $\mathrm{col}(A)$.

- Para resolver $Ax = b$ sobredeterminado ($m > n$), com rank(A) = n, **projetamos** b **sobre col**(A) **para que exista solução**. Dispondo de Q e R tais que $A = QR$ segue que

$$QRx^* = \Pi\, b \quad \Rightarrow \quad \left(\begin{array}{c|c} Q_1 & Q_2 \end{array} \right) \left(\frac{R_1}{0} \right) x^* \;=\; Q_1\, Q_1^T\, b$$

$$\Rightarrow \quad Q_1 R_1 x^* = Q_1 Q_1^T b$$

Multiplicando por Q_1^T, já que $Q_1^T Q_1 = I_{n\times n}$,

$$R_1\, x^* = Q_1^T\, b \quad \longrightarrow$$

```
Q1=Q(:,1:n);
R1=R(1:n,:);
x=R1\(Q1'*b)
```

- Resolver com fatoração QR é mais estável que resolver $A^T Ax^* = A^T b$.
- A construção de Q e R deve ser feita de maneira estável. O algoritmo básico de álgebra linear (ortogonalização de Gram-Schmidt) é instável. O **algoritmo de Householder** é preferível.
- Temos assim resolvido o problema de ajuste por mínimos quadrados quando rank$(A) = n$. O "erro" ou "resíduo" é $r = Ax^* - b = \Pi b - b = -Q_2 Q_2^T b$. Ele é ortogonal a col(A) e por tanto não pode ser melhorado.
- **Em Octave** `x=A\b` **fornece a solução de quadrados mínimos!** (não é necessário escrever `x=R1\(Q1'*b)`)
- O custo computacional é $\simeq 2mn^2 - 2n^3/3$. Se $m = n$ dá $4n^3/3$ (LU:$2n^3/3$).

Fatoração QR reduzida

Uma representação mais compacta para o caso $m > n$ é obtida eliminando as $m - n$ últimas colunas de Q (irrelevantes) e as últimas $m - n$ linhas de R (nulas).

Teorema: Seja $A \in \mathbb{R}^{m\times n}$, então existem $Q_1 \in \mathbb{R}^{m\times n}$, cujas colunas são ortonormais, e $R_1 \in \mathbb{R}^{n\times n}$, triangular superior com $r_{ii} \geq 0$ tais que

$$A = Q_1 R_1 \quad \longrightarrow \quad \texttt{[Q1 R1] = qr(A,0)}$$

Quando $\text{rank}(A) = n$ a matriz Q_1 é base de $\text{col}(A)$.
Quando $\text{rank}(A) < n$ existem *relações ocultas* entre as colunas, como descobrir?

Fatoração QR ordenada

Teorema: Seja $A \in \mathbb{R}^{m\times n}$, então existem $Q \in \mathbb{R}^{m\times m}$ ($Q_1 \in \mathbb{R}^{m\times n}$ na QR reduzida) cujas colunas são ortonormais, $R \in \mathbb{R}^{m\times n}$ ($R_1 \in \mathbb{R}^{n\times n}$ na QR reduzida), triangular superior com $r_{ii} \geq 0$, e uma matriz de permutação P tais que

$$AP = QR = Q_1\,R_1 \quad \longrightarrow$$

```
[Q R P] = qr(A)
[Q1 R1 P] = qr(A,0)
```

com $r_{ii} \geq r_{jj}$ se $i < j$.

Os primeiros $k = \text{rank}(A)$ valores diagonais de R (ou R_1) são positivos. Os $n - k$ restantes são zero. As primeiras k colunas de Q (ou Q_1) são base ortonormal de $\text{col}(A)$.

Trabalhando na representação reduzida, é

$$B = AP = Q_1 R_1 = \underbrace{\left(\begin{array}{c|c} \overline{Q}_1 & \overline{Q}_1^{\perp} \end{array} \right)}_{Q_1} \underbrace{\left(\begin{array}{c|c} \overline{R}_{11} & T \\ \hline 0 & 0 \end{array} \right)}_{R_1}$$

onde $\overline{Q}_1$ é de $m \times k$ com colunas ortonormais, $\overline{R}_{11}$ é triangular superior $(k \times k)$ com diagonal positiva, e T é de $k \times (n-k)$. Notar que $\overline{Q}_1^{\perp}$ é irrelevante.

Por inspeção,

$$\overline{B}_1 = (c_1(AP)| \ldots |c_k(AP)) = \overline{Q}_1 \overline{R}_{11}$$

e, para as colunas $j > k$,

$$c_j(AP) = \overline{Q}_1\, c_{j-k}(T) = \overline{B}_1 \overline{R}_{11}^{-1} c_{j-k}(T) = \overline{B}_1\, c_j(\overline{R}_{11}^{-1} T)$$

A coluna j de AP é combinação linear de $c_1(AP) \ldots c_k(AP)$.
Os coeficientes são a coluna j de $\overline{R}_{11}^{-1} T$.

```
> A=[[1 2 3 4]',[2 4 6 8]',[-1 1 -1 1]'];
> [q r p]=qr(A,0)
q =
  -0.182574   0.542451   0.818378
  -0.365148  -0.440741   0.259794
  -0.547723   0.610257  -0.510329
  -0.730297  -0.372935   0.048255
r =
  -10.95445   -0.36515   -5.47723
    0.00000   -1.96638   -0.00000
    0.00000    0.00000   -0.00000
p =
   2   3   1 (i.e. p=[0 1 0;0 0 1;1 0 0]')
> r11=r(1:2,1:2);q11=q(:,1:2);t=r(1:2,3);
```

```
> S=r11\t
S =
   5.0000e-01
   3.7604e-16
```

Isto nos indica que a **terceira coluna** de AP (que é a primeira coluna de A), é igual a 0.5 vezes a **primeira coluna** de AP (que é a segunda coluna de A).

Assim podemos “descobrir” relações lineares entre nossas colunas de dados.

Resumo:

Temos resolvido:

- Sistemas sobredeterminados ($m > n$).
- Ajustar uma coluna de A como uma combinação linear de outras colunas ou de funções delas.
- Determinar $\mathrm{rank}(A)$ e base de $\mathrm{col}(A)$.
- Descobrir relações lineares exatas entre colunas de A.

Ficam duas perguntas naturais pendentes:

- Como resolver $Ax = b$ se $\mathrm{rank}(A) < n$? (exercício opcional)
- Como descobrir relações lineares **aproximadas** entre colunas de dados? (para isto, SVD)