Ajuste de Curvas (Parte II)

Prof. Americo Cunha

Universidade do Estado do Rio de Janeiro – UERJ

americo.cunha@uerj.br

www.americocunha.org

@AmericoCunhaJr

@AmericoCunhaJr @AmericoCunhaJr

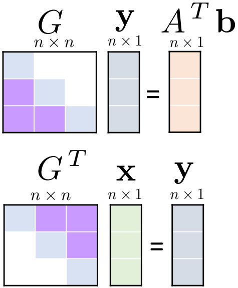
Resolvendo o problema de mínimos quadrados (computacionalmente)

Mínimos quadrados via fatoração Cholesky

$$\begin{cases} A^T A\mathbf{x} = A^T\mathbf{b} \\ A^T A = G\,G^T \end{cases} \iff \begin{cases} G\,\mathbf{y} = A^T\mathbf{b} \\ G^T\mathbf{x} = \mathbf{y} \end{cases}$$

Receita computacional

1. Construir $A^T A$ e $A^T\mathbf{b}$
2. Calcular $A^T A = G\,G^T$
3. Resolver $G\,\mathbf{y} = A^T\mathbf{b}$
4. Resolver $G^T\mathbf{x} = \mathbf{y}$



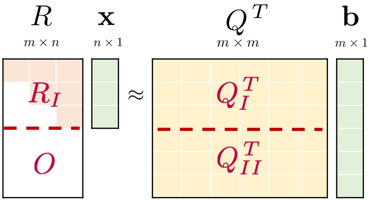
$$\texttt{flops}\,(\textbf{MQ Cholesky}) \sim m\,n^2 + \frac{1}{3}\,n^3$$

Mínimos quadrados via fatoração QR

$$\left\{\begin{array}{lcl} A^T A\,\mathbf{x} & = & A^T\,\mathbf{b} \\ A & = & Q\,R \end{array}\right. \iff \left\{\begin{array}{lcl} R_I\,\mathbf{x} & = & Q_I^{\,T}\,\mathbf{b} \end{array}\right.$$

Receita computacional

1. Calcular $A = Q\,R$
2. Calcular $Q_I^{\,T}\,\mathbf{b}$
3. Resolver $R_I\,\mathbf{x} = Q_I^{\,T}\,\mathbf{b}$



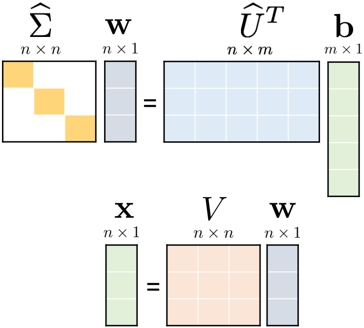

```
x = A \ b
```

"backslash" usa fatoração QR para sistemas retangulares

$$\texttt{flops}\,(\mathbf{MQ\ QR}) \sim 2\,m\,n^2 - \frac{2}{3}\,n^3 + n\,m + n^2$$

Mínimos quadrados via fatoração SVD

$$\begin{cases} A^T A\mathbf{x} = A^T\mathbf{b} \\ A = \widehat{U}\,\widehat{\Sigma}\,V^T \end{cases} \iff \begin{cases} \widehat{\Sigma}\,\mathbf{w} = \widehat{U}^T\mathbf{b} \\ \mathbf{x} = V\,\mathbf{w} \end{cases}$$



Receita computacional

1. Calcular $A = \widehat{U}\,\widehat{\Sigma}\,V^T$
2. Calcular $\widehat{U}^T\,\mathbf{b}$
3. Resolver $\widehat{\Sigma}\,\mathbf{w} = \widehat{U}^T\,\mathbf{b}$
4. Calcular $\mathbf{x} = V\,\mathbf{w}$

$$\texttt{flops}\,(\textbf{MQ SVD}) \sim 2\,m\,n^2 + 11\,n^3 + m^2 + n + n^2$$

Custo do problema de mínimos quadrados ($m \gg n$)

fatoração	flops	estabilidade	custo
Cholesky	$\sim m\,n^2 + \frac{1}{3}\,n^3$	$\star\star\star$	$1 \times$ \$
QR	$\sim 2\,m\,n^2 - \frac{2}{3}\,n^3$	$\star\star\star\star$	$2 \times$ \$
SVD	$\sim 2\,m\,n^2 + 11n^3$	$\star\star\star\star\star$	$2 \times$ \$

Podemos usar outra noção de erro?

Ajuste de curva robusto

Outras medidas de erro são possíveis (e pertinentes):

- Erro médio quadrático

$$E_2 = \left(\frac{1}{m} \sum_{k=1}^{m} \left(y_k - f(x_k) \right)^2 \right)^{1/2}$$

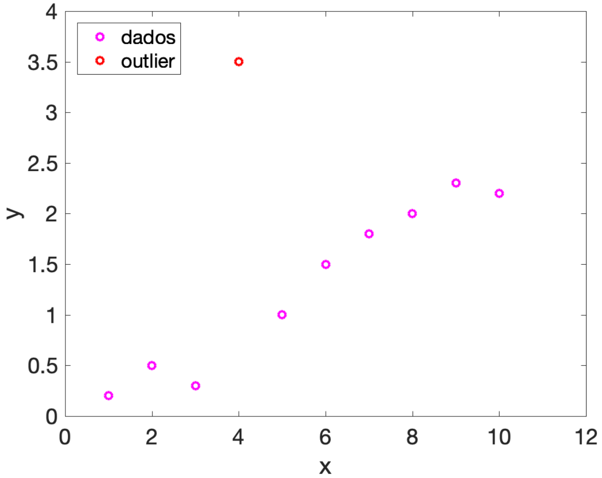
- Erro médio

$$E_1 = \frac{1}{m} \sum_{k=1}^{m} |y_k - f(x_k)|$$

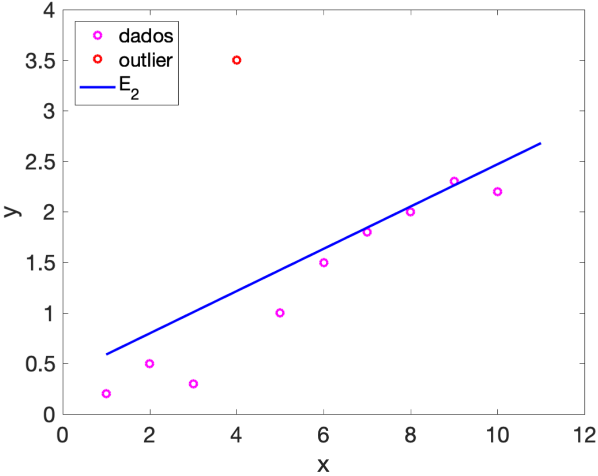
- Erro máximo

$$E_\infty = \max_{1 \leq k \leq m} |y_k - f(x_k)|$$

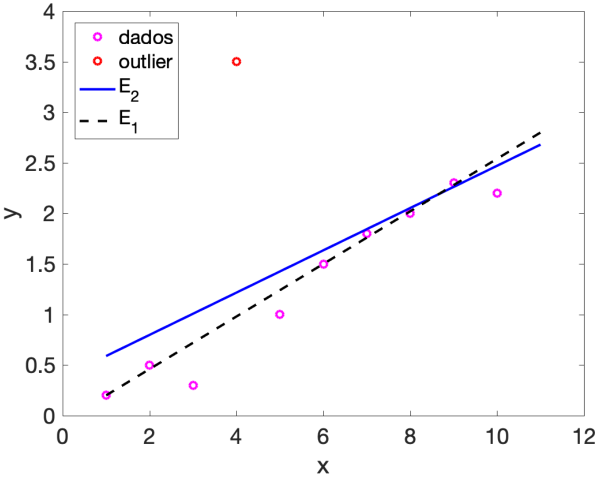
Ajuste de curva robusto



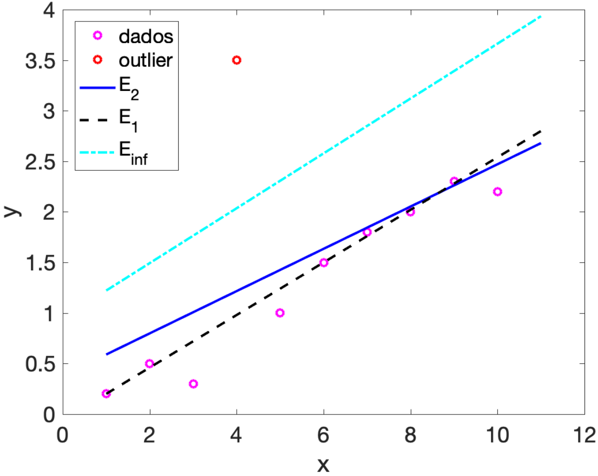
Ajuste de curva robusto



Ajuste de curva robusto



Ajuste de curva robusto

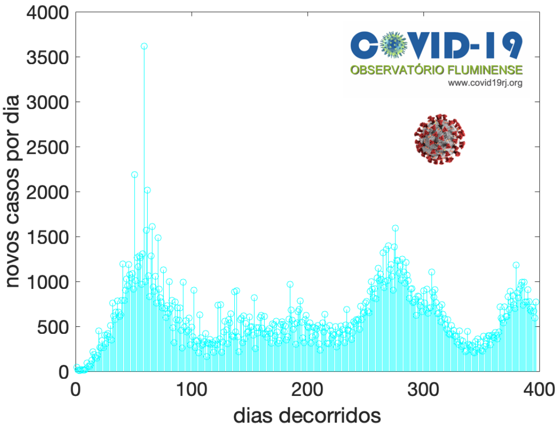


Ajuste de curva robusto

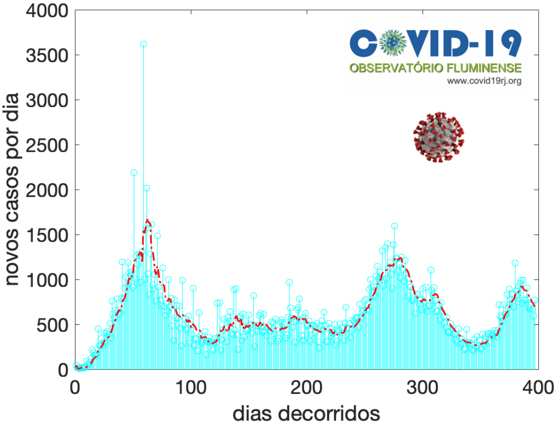
MainRegressionExample3.m

```
clc; clear; close all;

m = 10;
xd = [1 2 3 4 5 6 7 8 9 10];
yd = [0.2 0.5 0.3 3.5 1.0 1.5 1.8 2.0 2.3 2.2];

E_2   = @(p) sqrt(sum(abs(p(1)*xd+p(2)-yd).^2)/m);
E_1   = @(p)      sum(abs(p(1)*xd+p(2)-yd)   )/m;
E_inf = @(p)  max(abs(p(1)*xd+p(2)-yd));
p_2   = fminsearch(E_2  ,[1 1]);
p_1   = fminsearch(E_1  ,[1 1]);
p_inf = fminsearch(E_inf,[1 1]);

xfit  = 1:0.01:11;
y_2   = polyval(p_2  ,xfit);
y_1   = polyval(p_1  ,xfit);
y_inf = polyval(p_inf,xfit);

plot(xd,yd,'om',xd(4),yd(4),'or','LineWidth',2);
hold on
plot(xfit,y_2  ,'-b' ,'LineWidth',2);
plot(xfit,y_1  ,'--k','LineWidth',2);
plot(xfit,y_inf,'-.c','LineWidth',2);
hold off
xlabel('x'); ylabel('y');
set(gca,'FontSize',18);
legend('dados','outlier','E_2','E_1','E_{inf}')
```

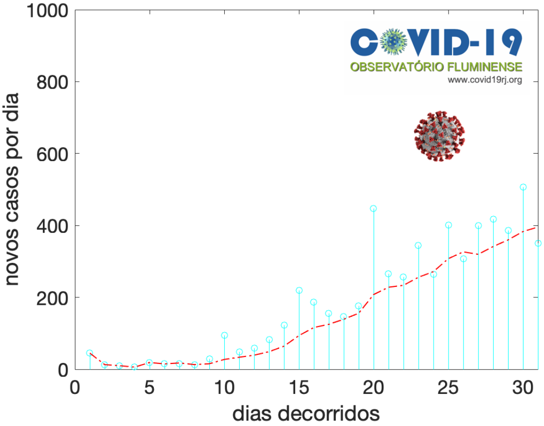
COVID-19 na cidade do Rio de Janeiro



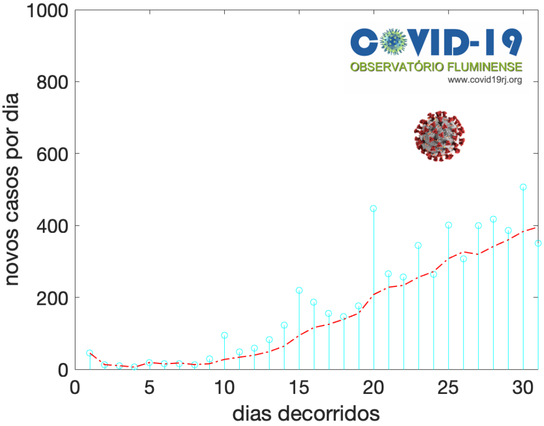
COVID-19 na cidade do Rio de Janeiro



COVID-19 na cidade do Rio de Janeiro



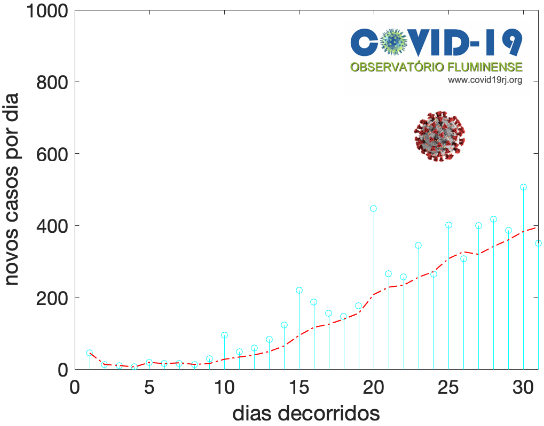
COVID-19 na cidade do Rio de Janeiro



(fase de expansão)

novos casos $\sim N \exp(\beta\, t)$

COVID-19 na cidade do Rio de Janeiro

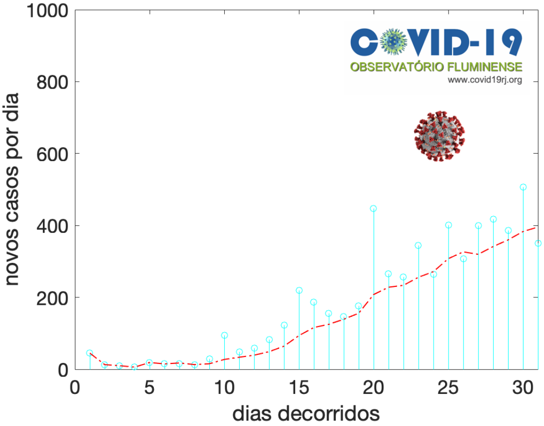


(fase de expansão)

novos casos $\sim N \exp(\beta\, t)$

$$y = N \exp(\beta\, t)$$

COVID-19 na cidade do Rio de Janeiro



(fase de expansão)

novos casos $\sim N \exp(\beta\, t)$

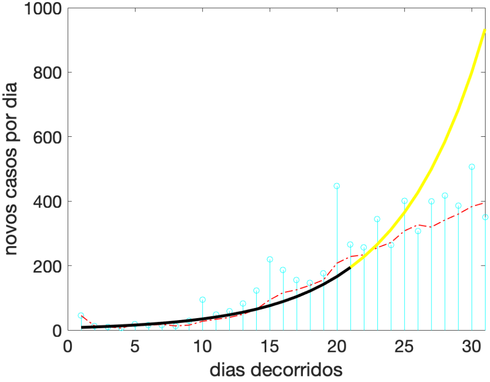
$$y = N \exp(\beta\, t) \implies \ln y = \beta\, t + \ln N$$

COVID-19 na cidade do Rio de Janeiro

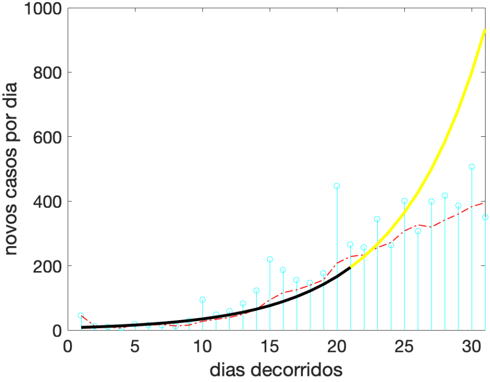
MainRegressionExample4.m

```
clc; clear; close all;

dados = load('covid19_casos_rio_2020_2021.csv');

mm7 = dados;
for j=7:length(dados)
    mm7(j) = sum(dados(j-6:j))/7;
end

xdata = 1:21; ydata = mm7(xdata)';
A     = [xdata; ones(size(xdata))]';
b     = log(ydata)';
x     = A\b;
yfit  = @(z) exp(x(2))*exp(x(1)*z);

stem(dados(1:31),'oc','LineWidth',0.5);
hold on
plot(1:31,  mm7(1:31),'-.r','LineWidth',1);
plot(yfit(1:31),'-y','LineWidth',3);
plot(yfit(1:21),'-k','LineWidth',3);
hold off
xlabel('dias decorridos'); ylabel('novos casos por dia');
set(gca,'FontSize',18);
```

COVID-19 na cidade do Rio de Janeiro



COVID-19 na cidade do Rio de Janeiro



Outra possibilidade, curva logística (via regressão não linear):

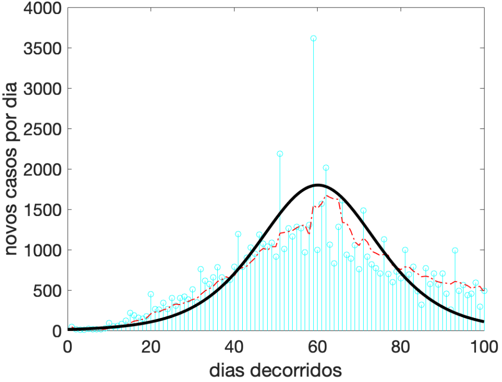
$$I(x) = \frac{\alpha\, L\, e^{-\alpha\,(t - t_m)}}{\left(1 + e^{-\alpha\,(t - t_m)}\right)^2}$$

COVID-19 na cidade do Rio de Janeiro

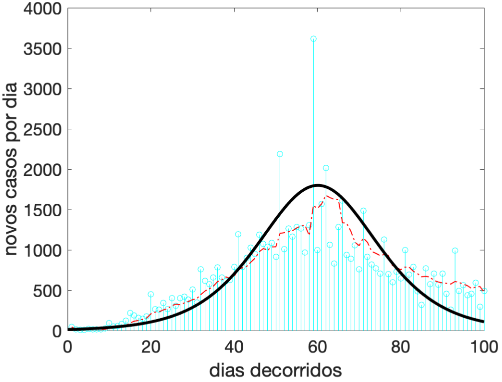
MainRegressionExample5.m

```
clc; clear; close all;

dados = load('covid19_casos_rio_2020_2021.csv');

mm7 = dados;
for j=7:length(dados)
    mm7(j) = sum(dados(j-6:j))/7;
end

xdata = 1:100; ydata = mm7(xdata)';
N     = @(x,p) p(1)*p(2)*exp(-p(1)*(x-p(3)));
D     = @(x,p) (1+exp(-p(1)*(x-p(3)))).^2;
I     = @(x,p) N(x,p)./D(x,p);
E_2   = @(p) sqrt(sum(I(xdata,p) - ydata).^2/length(ydata));
p_2   = fminsearch(E_2,[0.1 8e4 55]);
yfit  = I(0:0.1:100,p_2);

stem(dados(1:100),'oc','LineWidth',0.5);
hold on
plot(1:100    ,mm7(1:100),'-.r','LineWidth',1);
plot(0:0.1:100,yfit      ,'-k' ,'LineWidth',3);
hold off
xlabel('dias decorridos'); ylabel('novos casos por dia');
set(gca,'FontSize',18);
```

COVID-19 na cidade do Rio de Janeiro

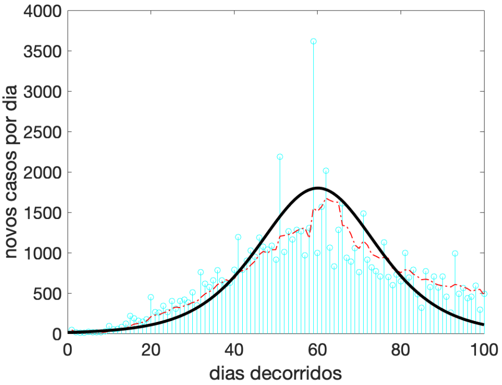


COVID-19 na cidade do Rio de Janeiro



E para ajustar várias ondas?

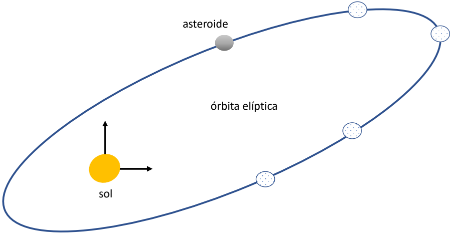
COVID-19 na cidade do Rio de Janeiro

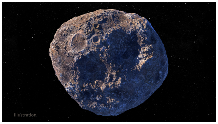


E para ajustar várias ondas?

Tente somar várias curvas logísticas!

Órbita de um asteroide





5 observações estão disponíveis

x (ua)	y (ua)
8,025	8,310
10,170	6,355
11,202	3,212
10,736	0,375
9,092	-2,267

ua = unidade astronômica

$$A\,x^2 + B\,x\,y + C\,y^2 + D\,x + E\,y + 1 = 0$$

Determine a equação da órbita!

Algumas bases de dados

registro civil
brasil.io
openDataSUS
dados.gov.br
ipea Instituto de Pesquisa Econômica Aplicada
IBGE Instituto Brasileiro de Geografia e Estatística
DataUERJ 2020
kaggle
Quandl
FiveThirtyEight
Socrata
aws
IBM Cloud
Google Cloud
Azure
Google Trends
Google Dataset Search
Our World in Data
EUROPEAN DATA PORTAL
DATA.GOV

Pegue alguns dados nessas bases e construa regressores!

Como citar esse material?

A. Cunha Jr, *Ajuste de Curvas*,
Universidade do Estado do Rio de Janeiro – UERJ, 2024.

@AmericoCunhaJr

@AmericoCunhaJr @AmericoCunhaJr