

# CC-226 Aula 03 - Análise de Regressão

Carlos Henrique Q. Forster - Instituto Tecnológico de Aeronáutica

2008

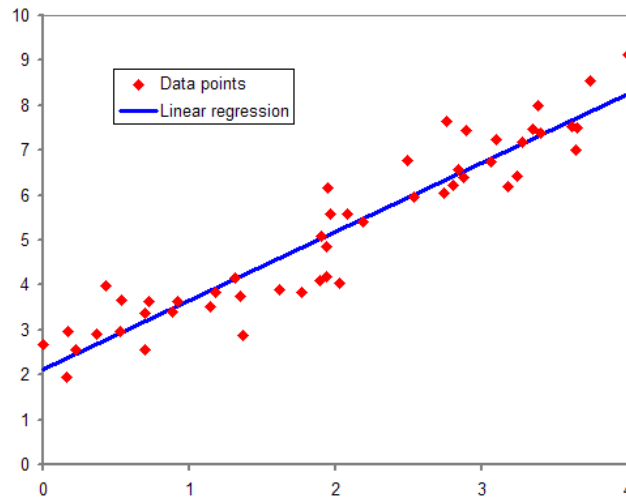
## 1 Regressão Linear Simples

### 1.1 Descrição do problema de regressão linear simples

Relação entre duas variáveis  $x$  e  $y$ .

$$y = \beta_0 + \beta_1 x$$

onde  $\beta_1$  é o coeficiente angular e  $\beta_0$  é o termo constante.



*Modelo de regressão linear simples:*

Desejamos estimar os parâmetros  $\beta_0$ ,  $\beta_1$  e  $\sigma^2$ .

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

onde  $\varepsilon$  é uma variável aleatória com média  $E(\varepsilon) = 0$ .

A variância  $Var(\varepsilon)$  é dada por  $\sigma^2$ .

$\varepsilon$ : desvio aleatório ou erro aleatório.

Dado um valor para  $x$  o valor esperado de  $Y$  é:

$$\begin{aligned} E(Y) &= E(\beta_0 + \beta_1 x + \varepsilon) = \beta_0 + \beta_1 x + E(\varepsilon) = \beta_0 + \beta_1 x \\ Var(Y) &= Var(\beta_0 + \beta_1 x + \varepsilon) = Var(\varepsilon) = \sigma^2 \end{aligned}$$

## 1.2 Estimação dos parâmetros de Regressão

$n$  pares observados  $(x_1, y_1), \dots, (x_n, y_n)$ .

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i : \varepsilon_i \text{ são independentes}$$

*Princípio dos mínimos quadrados*

Função a minimizar

$$f(b_0, b_1) = \sum_{i=1}^n [y_i - (b_0 + b_1 x_i)]^2$$

A solução da minimização é representada utilizando a notação:

$$(\hat{b}_0, \hat{b}_1) = \arg \max_{(b_0, b_1)} f(b_0, b_1)$$

Para minimizar, igualamos o gradiente de  $f$  com zero.

$$\nabla f = \mathbf{0}$$

$$\frac{\partial f}{\partial b_0} = \sum 2(y_i - b_0 - b_1 x_i)(-1) = 0$$

$$\frac{\partial f}{\partial b_1} = \sum 2(y_i - b_0 - b_1 x_i)(-x_i) = 0$$

É obtido o sistema de equações normais:

$$\begin{cases} nb_0 + (\sum x_i) b_1 = \sum y_i \\ (\sum x_i) b_0 + (\sum x_i^2) b_1 \end{cases}$$

A estimativa obtida é resultado desse sistema. O coeficiente angular (slope) estimado é

$$b_1 = \hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}$$

O termo constante (y-intercept) estimado é

$$b_0 = \hat{\beta}_0 = \frac{\sum y_i - \hat{\beta}_1 \sum x_i}{n} = \bar{y} - \hat{\beta}_1 \bar{x}$$

## 1.3 Avaliação da regressão

*Estimação de  $\sigma^2$*

Resíduos  $y_i - \hat{y}_i$  onde  $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ .

Soma dos quadrados dos erros (error sum of squares):

$$ESS = \sum (y_i - \hat{y}_i)^2 = \sum [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2$$

$$= \sum y_i^2 - \hat{\beta}_0 \sum y_i - \hat{\beta}_1 \sum x_i y_i$$

A estimativa de  $\sigma^2$  é

$$\hat{\sigma}^2 = s^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n - 2}$$

Notar a perda de dois graus de liberdade por  $\hat{\beta}_0$  e  $\hat{\beta}_1$  já estarem definidos.  
Soma dos quadrados totais (total sum of squares):

$$TSS = S_{yy} = \sum (y_i - \bar{y})^2$$

*Coefficiente de determinação (coeficiente de regressão de Pearson)*

$$r^2 = 1 - \frac{ESS}{TSS}$$

$r^2$  mede a proporção da variância que pode ser explicada pelo modelo de regressão linear simples. Quanto maior  $r^2$ , melhor adaptado ao modelo linear de regressão.

*Inferência sobre  $\beta_1$*

Supondo distribuição normal,  $\hat{\beta}_1$  é dada por uma combinação linear das variáveis  $y_i$  normalmente distribuídas (e, portanto, é uma variável aleatória normalmente distribuída).

$$E(\hat{\beta}_1) = \beta_1$$

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$$

*Correlação*

Coeficiente de correlação amostral

$$r = \frac{S_{xy}}{\sqrt{S_{xx}} \sqrt{S_{yy}}}$$

onde

$$S_{xy} = \sum_{i=1}^n x_i y_i - \frac{\sum x_i \sum y_i}{n}$$

Propriedades de  $r$

- não depende de qual variável é  $x$  e qual é  $y$ ;
- independe das unidades de medida;
- $-1 \leq r \leq 1$ ;
- $r = 1$  indica linha reta com coeficiente angular positivo;

- $r = -1$  indica linha reta com coeficiente angular negativo;
- $r^2$  é o coeficiente de determinação.

Regra prática

$0 \leq |r| \leq 0,5$  indica correlação fraca.

$0,8 \leq |r| \leq 1$  indica correlação forte.

$r^2 = 0,25$  indica que apenas 25% da variação de  $y$  observada seria explicada pelo modelo de regressão linear.

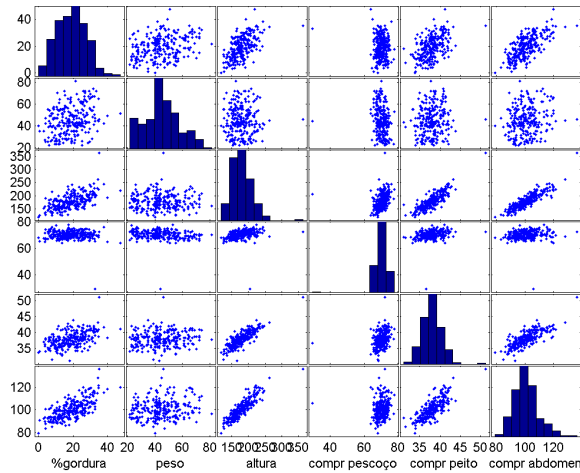
Coeficiente de correlação

$$\rho = \rho(X, Y) = \frac{Cov(X, Y)}{\sigma_x \sigma_y}$$

$$\hat{\rho} = R = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$$

## 2 Regressão Linear Múltipla

Exemplo do conjunto de dados "bodyfat" da statlib.



### 2.1 O modelo de regressão linear múltipla

Modelo

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m + \varepsilon$$

Onde

$$E(\varepsilon) = 0$$

e

$$Var(\varepsilon) = \sigma^2$$

$m$  é o número de parâmetros e  $n$  é o número de padrões de entrada.

Considerando  $n$  entradas da forma  $(y_i, x_{1i}, x_{2i}, \dots, x_{mi})$

Vamos considerar uma variável adicional  $X_0$  que resulta sempre em 1 e adicioná-la como campo nos padrões de entrada.

A função a minimizar pelo princípio dos mínimos quadrados é:

$$f = \sum_i (y_i - \beta_0 x_{0i} - \beta_1 x_{1i} \dots - \beta_m x_{mi})^2$$

Igualando suas derivadas parciais a zero:

$$\frac{f}{\beta_0} = -2 \sum_i x_{0i} (y_i - \beta_0 x_{0i} - \beta_1 x_{1i} \dots - \beta_m x_{mi}) = 0$$

$$\frac{f}{\beta_1} = -2 \sum_i x_{1i} (y_i - \beta_0 x_{0i} - \beta_1 x_{1i} \dots - \beta_m x_{mi}) = 0$$

...

Obtemos o sistema de equações normais.

O sistema de equações normais é da forma:

$$\begin{aligned} b_0 \sum x_{0i} x_{0i} &+ b_1 \sum x_{1i} x_{0i} &+ b_2 \sum x_{2i} x_{0i} &+ \dots &+ b_m \sum x_{mi} x_{0i} &= \sum y_i x_{0i} \\ b_0 \sum x_{0i} x_{1i} &+ b_1 \sum x_{1i} x_{1i} &+ b_2 \sum x_{1i} x_{2i} &+ \dots &+ b_m \sum x_{1i} x_{mi} &= \sum x_{1i} y_i \\ &\vdots &&&& \\ b_0 \sum x_{0i} x_{mi} &+ b_1 \sum x_{1i} x_{mi} &+ \dots &+ b_{m-1} \sum x_{m-1,i} x_{mi} &+ b_m \sum x_{mi} x_{mi} &= \sum x_{mi} y_i \end{aligned}$$

Se substituirmos a variável  $x_0$  por 1

$$\begin{aligned} b_0 n &+ b_1 \sum x_{1i} &+ b_2 \sum x_{2i} &+ \dots &+ b_m \sum x_{mi} &= \sum y_i \\ b_0 \sum x_{1i} &+ b_1 \sum x_{1i}^2 &+ b_2 \sum x_{1i} x_{2i} &+ \dots &+ b_m \sum x_{1i} x_{mi} &= \sum x_{1i} y_i \\ &\vdots &&&& \\ b_0 \sum x_{mi} &+ b_1 \sum x_{1i} x_{mi} &+ \dots &+ b_{m-1} \sum x_{m-1,i} x_{mi} &+ b_m \sum x_{mi}^2 &= \sum x_{mi} y_i \end{aligned}$$

## 2.2 Solução

Vamos escrever o problema de forma vetorial. O vetor  $\mathbf{x}$  é aumentado com um campo de valor 1.

$$\mathbf{x}_i = (1, x_{1i}, x_{2i}, x_{3i}, \dots, x_{mi})$$

O vetor  $\beta$  a ser determinado.

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{pmatrix}$$

Cada estimativa de saída  $\hat{y}_i$  é obtida com

$$\hat{y}_i = \sum_j \beta_j x_{ji} = \mathbf{x}_i \beta$$

Colocamos na forma matricial para todas entradas de  $i = 1, \dots, n$ :

$$X = \begin{pmatrix} 1 & x_{11} & x_{21} & x_{31} & \dots & x_{m1} \\ 1 & x_{12} & x_{22} & x_{j2} & \dots & x_{m2} \\ \vdots & & & & & \end{pmatrix}$$

O vetor-coluna  $\mathbf{y}$  contém todas variáveis de saída da amostra:

$$Y = \mathbf{y} = \begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

O sistema a ser resolvido (sub ou super determinado) é

$$X_{(n \times m+1)} \beta_{(m+1 \times 1)} = Y_{(n \times 1)}$$

O sistema de equações normais é

$$X^T X \beta = X^T Y$$

Assim, obtemos  $\beta$ :

$$\beta = (X^T X)^{-1} X^T Y$$

A matriz  $X^+ = (X^T X)^{-1} X^T$  é a matriz pseudo-inversa de Moore-Penrose de  $X$ .

Assim, a estimativa  $\hat{y}$  para as entradas  $x_j$  se escreve como:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m$$

*Avaliação da regressão múltipla*

A variância estimada para a estimativa  $\hat{y}$  é

$$\hat{\sigma}_\beta^2 = \frac{SQE}{n - (m + 1)}$$

Notar que houve perda de  $m + 1$  graus de liberdade

O coeficiente de determinação  $R^2$  deve ser ajustado para a regressão múltipla:

$$R_{ajustado}^2 = \frac{(n - 1)R^2 - m}{n - (m + 1)}$$

*Exemplo*

Considere duas variáveis  $x$  e  $z$  das quais queremos inferir  $y$  indiretamente. A partir de um conjunto de  $n$  padrões  $(x_i, z_i, y_i)$  estimamos os parâmetros da regressão.

O modelo do erro é

$$y_i = \beta_0 + x_i\beta_1 + z_i\beta_2 + \varepsilon_i$$

A função a minimizar (princípio dos mínimos quadrados):

$$f = \sum_i (y_i - \beta_0 - x_i\beta_1 - z_i\beta_2)^2$$

Igualando as derivadas parciais a zero:

$$\frac{\partial f}{\partial \beta_0} = -2 \sum_i (y_i - \beta_0 - x_i\beta_1 - z_i\beta_2) = 0$$

$$\frac{\partial f}{\partial \beta_1} = -2 \sum_i x_i (y_i - \beta_0 - x_i\beta_1 - z_i\beta_2) = 0$$

$$\frac{\partial f}{\partial \beta_2} = -2 \sum_i z_i (y_i - \beta_0 - x_i\beta_1 - z_i\beta_2) = 0$$

Na forma matricial:

$$\begin{bmatrix} 1 & \dots \\ x_1 & \dots \\ z_1 & \dots \end{bmatrix}_{3 \times n} \begin{bmatrix} 1 & x_1 & z_1 \\ \vdots & \vdots & \vdots \end{bmatrix}_{n \times 3} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}_{3 \times 1} = \begin{bmatrix} 1 & \dots \\ x_1 & \dots \\ z_1 & \dots \end{bmatrix}_{3 \times n} \begin{bmatrix} y_0 \\ y_1 \\ \vdots \end{bmatrix}_{n \times 1}$$

Multiplicando as matrizes:

$$\begin{bmatrix} 1 & \sum x_i & \sum z_i \\ \sum x_i & \sum (x_i)^2 & \sum x_i z_i \\ \sum z_i & \sum x_i z_i & \sum (z_i)^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum y_i x_i \\ \sum y_i z_i \end{bmatrix}$$

A matrix  $X^T X$  é simétrica e não-negativa definida, o que permite a decomposição de Cholesky ou a decomposição de auto-valores ou valores singulares (SVD). A pseudo-inversa do MATLAB utiliza a SVD.

Lembrando a decomposição de Cholesky:

$$LL^T = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{12} & l_{22} & 0 \\ l_{13} & l_{23} & l_{33} \end{bmatrix} \begin{bmatrix} l_{11} & l_{12} & l_{13} \\ 0 & l_{22} & l_{23} \\ 0 & 0 & l_{33} \end{bmatrix} = \begin{bmatrix} l_{11}^2 & l_{11} l_{12} & l_{11} l_{13} \\ l_{11} l_{12} & l_{12}^2 + l_{22}^2 & l_{12} l_{13} + l_{22} l_{23} \\ l_{11} l_{13} & l_{12} l_{13} + l_{22} l_{23} & l_{13}^2 + l_{23}^2 + l_{33}^2 \end{bmatrix}$$

Basta então igualar cada elemento dessa matriz com a matriz a se decompor e determinar um a um os coeficientes  $l_{ij}$ . Como a matriz  $L$  é triangular, é muito fácil resolver o sistema linear por substituição.

Na prática:

Construo a matriz  $X$ :

$$X = \begin{pmatrix} 1 & x_1 & z_1 \\ 1 & x_2 & z_2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & z_n \end{pmatrix}$$

Construo a matriz  $X^T X$ . Construo o vetor-coluna  $X^T Y$ . Decomponho  $X^T X$  em  $LL^T$  pela decomposição de Cholesky. Resolvo  $LL^T \beta = X^T Y$  por substituição e retro-substituição. Obtenho os parâmetros do vetor-coluna  $\beta$ .

Outra opção, no MATLAB uso `beta=pinv(X)*Y`.

### 2.3 Regressão múltipla para explorar relações entre variáveis

O modelo é utilizado mesmo para variáveis  $x_i$  dependentes permitindo previsores quadráticos e análise de interações de variáveis.

Exemplos de modelos:

Modelo de segunda ordem sem interação:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \varepsilon$$

Modelo de primeira ordem com interação:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

Modelo de segunda ordem completo:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \varepsilon$$

### 2.4 Complexidade

A solução do sistema linear de  $m$  variáveis e  $m$  equações é  $O(m^3)$ .

Se tenho  $m$  variáveis e busco as interações duas a duas, então terei  $O(m^2)$  variáveis e, portanto, a solução do sistema será  $O(m^6)$ .

## 3 Regressão Polinomial

Exemplo (Bishop): dados obtidos de

$$h(x) = \sin(2\pi x) + \varepsilon$$

Vamos supor  $n = 10$  observações  $(x_i, t_i)$ .

Chamamos o grau do polinômio de  $m$ . A forma do polinômio é

$$y(x, \mathbf{w}) = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_m x^m = \sum_{j=0}^M \beta_j x^j$$



Determinar os valores dos coeficientes (polynomial fitting).

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_M \end{pmatrix}$$

Função de erro a minimizar

$$f(\beta) = \frac{1}{2} \sum_{n=1}^N [y(x_n, \beta) - t_n]^2$$

Encontrar  $\beta^*$  para o qual  $f$  é mínimo.  
Igualar o gradiente de  $f$  a zero.

$$\frac{\partial f}{\partial \beta_0} = 0, \frac{\partial f}{\partial \beta_1} = 0, \dots, \frac{\partial f}{\partial \beta_m} = 0$$

É obtido o sistema de equações normais (como fizemos anteriormente).

$$\begin{array}{rcccccccc} \beta_0 n & + & \beta_1 \sum x_i & + & \beta_2 \sum x_i^2 & + & \dots & + & \beta_m \sum x_i^m & = & \sum y_i \\ \beta_0 \sum x_i & + & \beta_1 \sum x_i^2 & + & \beta_2 \sum x_i^3 & + & \dots & + & \beta_m \sum x_i^{m+1} & = & \sum x_i y_i \\ & & \vdots & & & & & & & & \\ \beta_0 \sum x_i^m & + & \beta_1 \sum x_i^{m+1} & + & \dots & + & \dots & + & \beta_m \sum x_i^{2m} & = & \sum x_i^m y_i \end{array}$$

É o mesmo que a regressão linear múltipla onde cada variável  $x_j = x^j$ .

*Avaliação da regressão polinomial*

$$ESS = \sum (y_i - \hat{y}_i)^2$$

$$TSS = \sum (y_i - \bar{y})^2$$

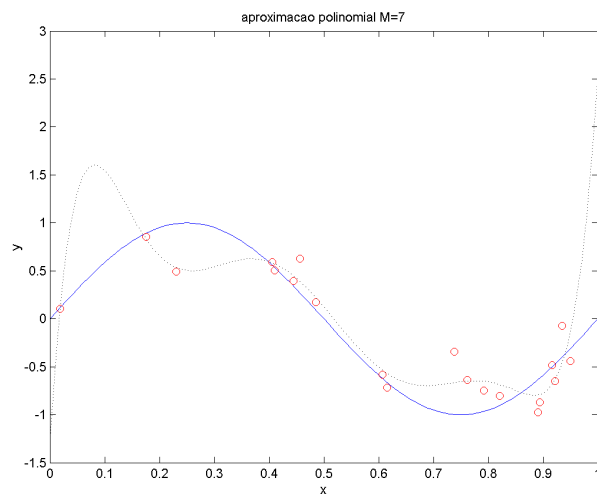
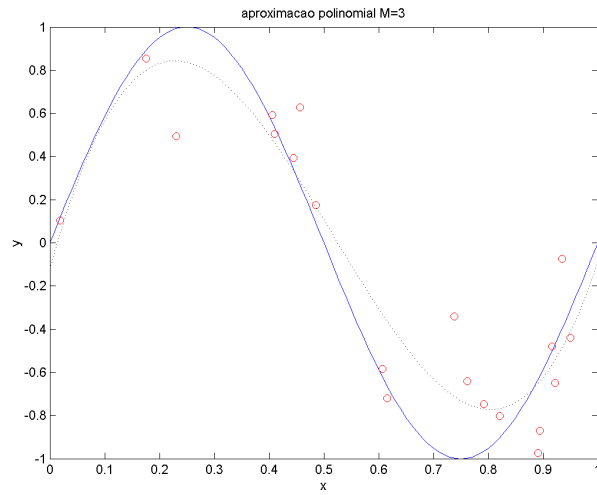
$$\hat{\sigma}^2 = s^2 = \frac{ESS}{n - (m + 1)}$$

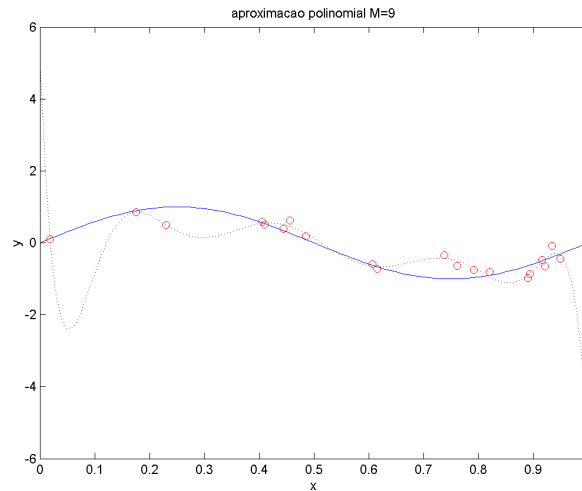
$$R^2 = 1 - \frac{ESS}{TSS}$$

$$R_{ajustado}^2 = 1 - \frac{n - 1}{n - (m + 1)} \frac{ESS}{TSS}$$

*Escolha do grau do polinômio*

Como escolher o valor para  $m$  Exemplo: no nosso exemplo,  $m = 0$  e  $m = 1$  são ruins.  $m = 3$  é bom.  $m = 9$  modela o ruído (overfitting).





Manter  $M$ , mas aumentar  $N$  melhora o problema de overfitting.  
 Treinar com um conjunto de 10, mas testar com um conjunto de 100: é boa a generalização?

Critério para teste: erro root-mean-square.

$$E_{RMS} = \sqrt{\frac{2f(\beta^*)}{N}}$$

### 3.1 Validação cruzada.

Validação cruzada S-fold:

1. Particionar o conjunto de amostras em  $S$  conjuntos.
2. Repito para cada um dos conjuntos:
  - (a) treinar com os demais  $S-1$  conjuntos.
  - (b) testar com o conjunto selecionado (Erro RMS).

## 4 Aproximação regularizada

Manter o número de parâmetros  $M$ , independente do número de amostras  $N$ .

Termo penalizando a função de erro para manter coeficientes pequenos.

$$\tilde{f}(\beta) = \frac{1}{2} \sum_{n=1}^N [y(x_n, \beta) - t_n]^2 + \frac{\lambda}{2} \|\beta\|^2$$

onde  $\|\beta\|^2 = \beta^T \beta = \beta_0^2 + \dots + \beta_M^2$ .

onde  $\lambda$  pesa a importância da regularidade da curva em relação ao erro contra os valores amostrais.

Termos em Estatística: "shrinkage" e "ridge regression".

A solução ainda pode ser obtida na forma fechada. Na forma matricial, a regularização que vimos é

$$(X^T X + \lambda I)\beta = X^T Y$$

A solução é então:

$$\beta = (X^T X + \lambda I)^{-1} X^T Y$$

Uma forma mais geral de regularização utiliza a matriz de Tikhonov  $\Gamma$ :

$$\beta = (X^T X + \lambda \Gamma^T \Gamma)^{-1} X^T Y$$

Para sintonizar automaticamente  $M$  e  $\lambda$ , em geral, é desperdício utilizar dados adicionais. Mesmo assim, esses parâmetros podem ser sintonizados utilizando validação cruzada. Entretanto, há necessidade de um conjunto adicional para validação do resultado.

#### 4.1 Código MATLAB

```

%% obter conjunto para treinamento
tam_trein=20 % numero de valores para treinamento
rand('state',0); randn('state',0);
x=rand(tam_trein,1); % 10 valores de x entre 0 e 1
h=sin(2*pi*x)+0.3*randn(tam_trein,1);

%% obter conjunto para testes
xt=0:0.01:1;
ht=sin(2*pi*xt);

%% plotar o original e as amostras
figure;
plot(xt,ht,'b-',x,h,'ro');
xlabel('x');
ylabel('y');
title('curva original e amostras com ruido');

%% obter aproximacoes polinomiais
for M=[1, 2, 3, 5, 7, 9];
x_mat=[];
for grau=0:M
    x_mat=[x_mat,x.^grau];
end
x_mat;

beta=pinv(x_mat)*h

%%plotar a aproximacao

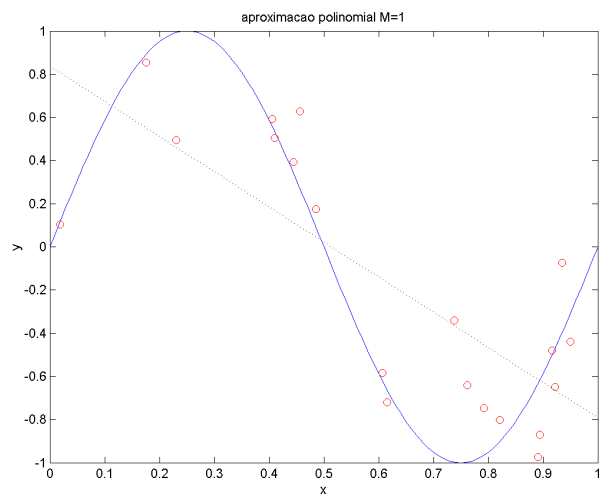
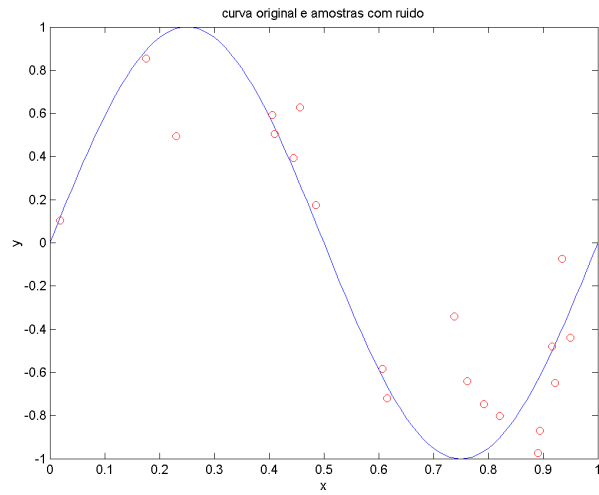
```

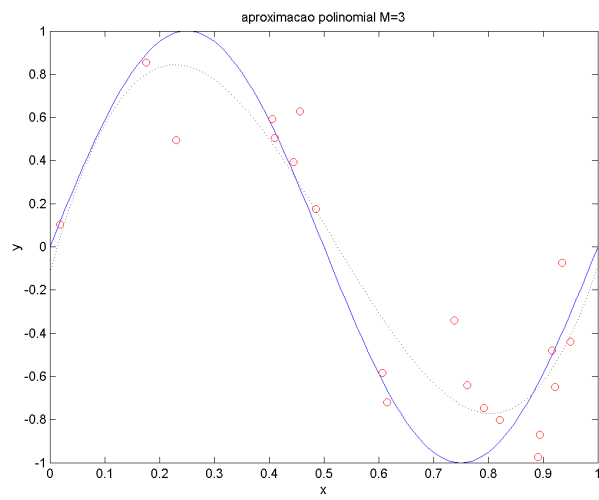
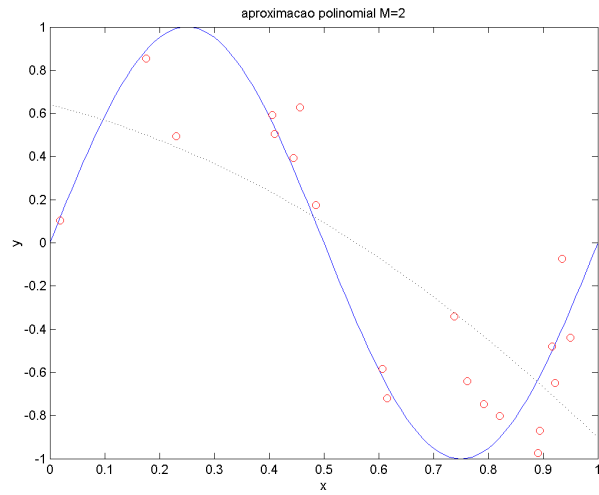
```
poly=flipud(beta);
p=polyval(poly,xt);
figure;
plot(xt,ht,'b-',x,h,'ro',xt,p,'g:');
xlabel('x');
ylabel('y');
title(['aproximacao polinomial M=',num2str(M)]);
end

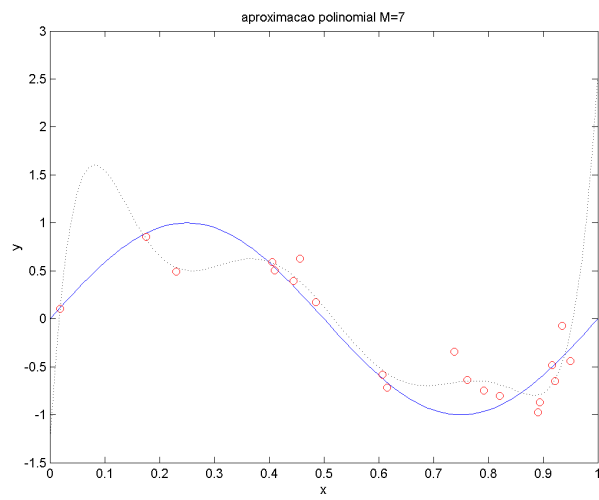
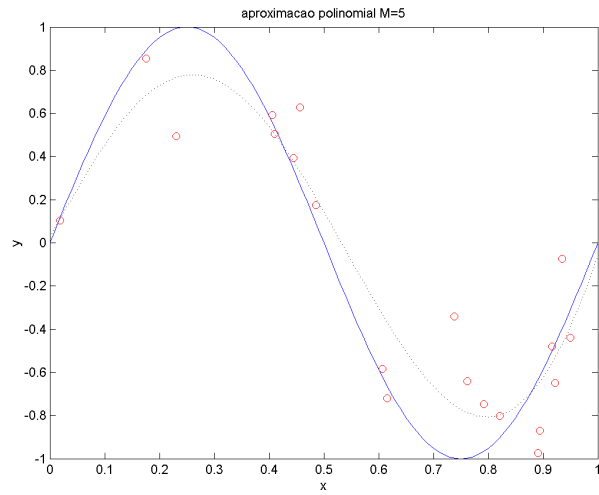
%% obter aproximacao regularizada
M=9;
lambda=0.001;
x_mat=[];
for grau=0:M
    x_mat=[x_mat,x.^grau];
end

xtx=x_mat'*x_mat;
beta=inv(xtx+lambda*eye(size(xtx)))*(x_mat'*h);

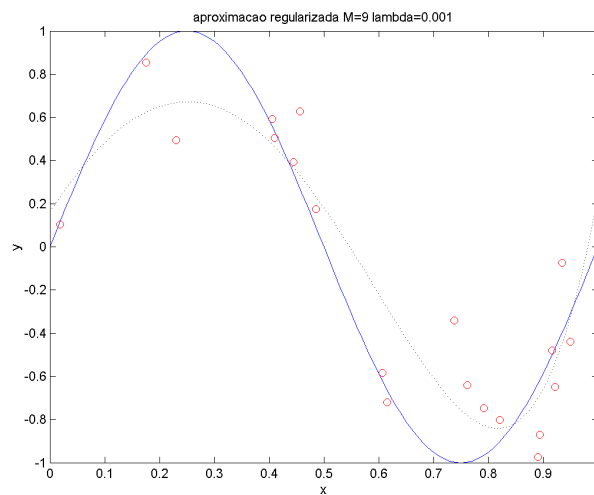
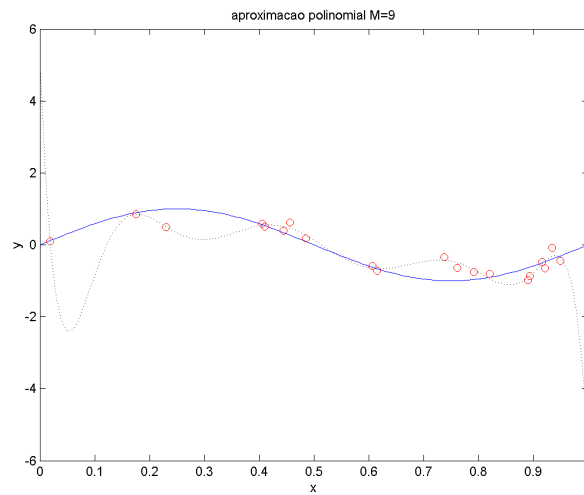
%%plotar aproximacao
poly=flipud(beta);
p=polyval(poly,xt);
figure;
plot(xt,ht,'b-',x,h,'ro',xt,p,'g:');
xlabel('x');
ylabel('y');
title(['aproximacao regularizada M=',num2str(M),' lambda=',num2str(lambda)]);
```











## 5 Regressão não-linear com bases funcionais

### 5.1 Radial Basis Function

A regressão linear utilizando bases de funções radiais tem forma:

$$\hat{y} = \beta_1 \phi_1(x) + \beta_2 \phi_2(x) + \beta_3 \phi_3(x)$$

A dimensão de  $x$  pode ser maior que 1.

As funções  $\phi_j$  tem a forma:

$$\phi_j = KF(|x - c_j|/KW)$$

KF é uma função de Kernel (por exemplo, uma Gaussiana) e KW é a largura do Kernel.  $c_j$  são coordenadas de centro para cada função.

Na prática,  $c_j$  e KW ou  $KW_j$  são mantidos constantes e só se estimam os valores de  $\beta$ .

Define-se uma matriz  $Z$  baseada na computação das funções  $\phi_j$ .

$$Z : z_{ji} = KF(|x_i - c_j|/KW)$$

A solução para os betas é:

$$\beta = (Z^T Z)^{-1} (Z^T Y)$$

## 5.2 RBF com ajuste não-linear

Utilizar método da descida na direção do Gradiente para ajustar  $\beta_j$ ,  $c_j$  e  $KW_j$ .

## 6 Heteroscedasticidade

É como se chama quando a variância do ruído não é constante para  $x$ .

Se temos entradas na forma  $(x_i, y_i, \sigma_i^2)$ , então, o problema de minimização é

$$\arg \min_{\beta} \sum \frac{(y_i - \beta x_i)^2}{\sigma_i^2}$$

Trata-se portanto de uma regressão com pesos  $\frac{1}{\sigma^2}$ .