



UNIVERSIDADE FEDERAL DO CEARÁ
CAMPUS SOBRAL

ENGENHARIA DA COMPUTAÇÃO
INTELIGÊNCIA COMPUTACIONAL

DENILSON GOMES VAZ DA SILVA

IMPLEMENTAÇÃO DOS MODELOS DE REGRESSÃO
(POLINOMIAL E LINEAR)

SOBRAL
2018

1. INTRODUÇÃO.....	2
2. IMPLEMENTAÇÃO.....	3
2.1 Regressão Polinomial.....	3
2.2 Regressão Linear.....	4
3. MÉTODOS DE AVALIAÇÃO DOS MODELOS.....	6
4. RESULTADOS.....	7
4.1 Regressão Polinomial.....	7
4.2 Regressão Linear.....	9

1. INTRODUÇÃO

A regressão múltipla é uma técnica que pode relacionar uma variável de saída a uma ou mais variáveis de entrada. Dessa forma, pode-se aferir o comportamento da variável dependente.

A regressão múltipla usa variáveis de entrada e de saída conhecidas para poder desvendar o comportamento da variável dependente (Saída) em relação a variável independente (Entrada), ou seja, você não precisa ter em mãos a função da variável dependente, basta aplicar regressão múltipla com as variáveis de entrada e de saída para obter uma aproximação dessa função.

No modelo de regressão múltipla polinomial a saída estimada é aproximada por um polinômio de grau n , como pode ser visto na equação 1.

$$\tilde{y} = a_0 + a_1x + a_2x^2 + a_3x^3 + \dots + a_nx^n + E \quad \text{Eq.(1)}$$

Já no modelo da regressão múltipla linear a saída estimada é aproximada por uma equação linear, como vemos na equação 2.

$$\tilde{y} = a_0 + a_1x + a_2x + a_3x + \dots + a_nx + E \quad \text{Eq.(2)}$$

A diferença entre a regressão polinomial e linear é exatamente a forma de representar a saída estimada, enquanto na polinomial é um polinômio, no linear é uma equação linear.

2. IMPLEMENTAÇÃO

2.1 Regressão Polinomial

Inicialmente aplicamos a regressão polinomial nos dados do arquivo “aerogerador.dat”. O arquivo “aerogerador.dat” contém os valores da variável de entrada (velocidade do vento em m/s) e da variável de saída (potência gerada em kWatts).

Para começar a regressão polinomial, nós perguntamos ao usuário o grau desejado para o polinômio (entre 2 e 6). Depois disso carregamos os dados do arquivo “aerogerador.dat” e atribuímos a variável de entrada para x e a variável de saída para y. Em seguida, construímos a matriz X com os polinômios de acordo com o grau escolhido. Este trecho do código pode ser visto na figura 1.

```
11      %recebe o grau do polinomio a ser gerado
12 -    grau = input('Digite o grau do polinômio (entre 2 e 6): ');
13
14 -    load aerogerador.dat %carrega o arquivo
15
16 -    x = aerogerador(:,1); %x recebe a coluna da variavel independente
17 -    y = aerogerador(:,2); %y recebe a coluna da variavel dependente
18
19 -    [tamanho,~] = size(x); %recebe o tamanho de entradas
20
21      %X é a matriz dos polinomio
22 -    X(:,1) = ones(tamanho,1); %para cada entrada atribuímos o valor 1
23      %na primeira coluna
24
25      %Criamos as colunas x, x^2, x^3 ,...,x^grau
26 -    for i = 1:grau
27 -        X(:,i+1) = aerogerador(:,1).^(i); %atribuímos a entrada elevada
28 -        %ao devido expoente
29 -    end
```

Figura 1. Trecho 1 da Regressão Polinomial.

Fonte: Próprio Autor

Após construir a matriz X, podemos obter o vetor “a” com as constantes que multiplicam o fator da variável independente. Daí então, já podemos construir a nossa saída. A construção da saída obtida por meio da regressão múltipla polinomial pode ser vista figura 2.

```

31 %operação com as matrizes para encontrarmos as constantes que multiplicam x
32 - a = ((X'*X)^(-1))*X'*y; %a contem as constantes do polinomio
33
34 %yout é a saída polinomial obtida por regressão
35 - yout = a(1); %yout recebe o valor da primeira constante
36
37 %yout é acrescido das outras componentes
38 - for j = 1:grau
39 -     yout = a(j+1)*x.^(j) + yout; %yout é acrescido da componente x^j
40 - end
41 %yout é o polinomio obtido por regressao que relaciona y e x.

```

Figura 2. Trecho 2 da Regressão Polinomial.

Fonte: Próprio Autor

Já com a saída estimada obtida, é hora de plotar os gráficos (saída do arquivo aerogerador.bat e saída obtida por regressão polinomial), verificar os seus comportamento em relação a variável independente e avaliar sua qualidade pelas métricas R^2 e R^2_{aj} . Os gráficos em questão podem ser vistos na seção **Resultados**. O trecho do código que exibe os gráficos e calcula o valor para R^2 e R^2_{aj} pode ser visto na figura 3.

```

43 - plot(x,y,'ko'); %plota as entradas x e as saidas y do arquivo aerogerador.dat
44 - hold on; %mantem o grafico x vs y
45 - plot(x,yout,'r'); %plota as entradas x e o polinomio obtido por regressao
46 - ylabel('Potência gerada (KW)'); %titulo eixo vertical
47 - xlabel('Velocidade do vento (m/s)'); %titulo eixo horizontal
48 %legend('Saidas do arquivo', 'Polinomio obtido por regressao', 'tl')
49
50 %cálculo de R^2 e R^2ajustado
51 - r1 = R(y,yout); %r1 recebe o coeficiente de determinação
52 - r2 = Raj(y,yout,X); %r2 recebe o coeficiente ajustado
53 - em = erroMedio(y,yout); %em recebe o erro medio
54
55 - str = ['Coeficiente de Determinação R^2: ' num2str(r1)];
56 - disp(str); %exibe a mensagem acima com o valor de r1
57
58 - str = ['Coeficiente de Determinação r^2 Ajustado: ' num2str(r2)];
59 - disp(str); %exibe a mensagem acima com o valor de r2
60 - str = ['Erro medio: ' num2str(em)];
61 - disp(str); %exibe a mensagem acima com o valor de em

```

Figura 3. Trecho para plotar e avaliar a saída obtida por regressão polinomial.

Fonte: Próprio Autor

2.2 Regressão Linear

A regressão múltipla linear foi aplicada na base de dados D, onde as colunas 1 e 2 são variáveis independentes (Entradas) e a coluna 3 é a variável dependente (Saída).

Com as duas variáveis de entradas montamos a matriz X. Este trecho do código pode ser visto na figura 4.

```

9      %base de dados D
10 -   D =[122 139 0.115;
11       114 126 0.120;
12       086 090 0.105;
13       134 144 0.090;
14       146 163 0.100;
15       107 136 0.120;
16       068 061 0.105;
17       117 062 0.080;
18       071 041 0.100;
19       098 120 0.115];
20
21 -   x1 = D(:,1); %variavel independente x1
22 -   x2 = D(:,2); %variavel independente x2
23 -   y = D(:,3); %variavel dependente y
24
25     %matriz X
26 -   X(:,1) = x1;
27 -   X(:,2) = x2;

```

Figura 4. Trecho para criação da base e da matriz X.

Fonte: Próprio Autor

Com a matriz X em mãos podemos calcular o vetor B, que contém as constantes que multiplicam as variáveis de entrada. Após obter estas constantes basta multiplicar a matriz X por B e obtemos a saída *yout*. Esse trecho do código é visto na figura 5.

```

29 -   B = ((X'*X)^(-1))*X'*y; %B contem as constantes
30 -   yout = X*B; %saida pela regressao multipla

```

Figura 5. Trecho para obtenção da saída estimada.

Fonte: Próprio Autor

Depois de obter a saída, é hora de plotar o gráfico com a saída da base de dados D e a saída obtida pela regressão múltipla linear *yout*. Vamos também avaliar o método pela métrica R^2 . Os gráficos e os valores de R^2 são exibidos na seção Resultados. O trecho do código que plota as saídas e calcula o valor de R^2 é visto na figura 6.

```

32 -   plot3(x1,x2,y,'k'); %plota as entradas x1 e x2 para as saidas y da base
33 -   hold on; %mantem o grafico
34 -   grid on; %habilita as linhas de grade
35 -   plot3(x1,x2,yout,'r'); %plota as entradas x1 e x2 para a saida obtida por regressao
36 -   xlabel('X1'); %eixo x1
37 -   ylabel('X2'); %eixo x2
38 -   zlabel('Yout'); %eixo yout
39 -   legend('Saidas da Base', 'Saida da Regressao Multipla', 'nl')
40
41     %cálculo de R^2
42 -   r = R(y,yout); %r recebe o coeficiente de determinação
43 -   EM = erroMedio(y,yout); %EM recebe o erro medio
44
45 -   str = ['Coeficiente de Determinação R^2: ' num2str(r)];
46 -   disp(str); %exibe a mensagem acima com o valor de r
47 -   str = ['Erro Medio: ' num2str(EM)];
48 -   disp(str); %exibe a mensagem acima com o valor de EM

```

Figura 6. Trecho para plotagem das saídas e avaliação do método.

Fonte: Próprio Autor

3. MÉTODOS DE AVALIAÇÃO DOS MODELOS

Para avaliar os modelos desenvolvidos, foram criadas 3 funções: a função “*R*” que calcula o valor do coeficiente de determinação pela métrica R^2 , a função “*Raj*” que calcula o valor do coeficiente de determinação pela métrica R^2_{aj} (R^2 ajustado) e uma função “*erroMedio*” que calcula o erro médio das aproximações obtidas por regressão múltipla.

A função *R* e *Raj* pode ser vistas nas figuras 7 e 8, respectivamente.

```
8 function coef = R(Y,Yout)
9 [tamanho,~] = size(Y); %tamanho recebe o numero de saidas
10
11 ymed = sum(Y)/tamanho; %ymed recebe o valor medio das saidas Y
12
13 SQe = 0;
14 for k = 1:tamanho
15     SQe = SQe + (Y(k) - Yout(k))^2;
16 end
17 %SQe recebe o somatorio dos quadrados das diferenças entre y e yout
18
19 Syy=0;
20 for w = 1:tamanho
21     Syy = Syy + (Y(w) - ymed)^2;
22 end
23 %Syy recebe o somatorio dos quadrados das diferenças entre y e ymed
24
25 coef = 1 - SQe/Syy; %coef vale o coeficiente de determinação
```

Figura 7. Função *R* (retorna coeficiente pela métrica R^2).

Fonte: Próprio Autor

```
8 function coef = Raj(Y,Yout,poli)
9
10 [~,grau] = size(poli); %grau do polinimio
11 [tamanho,~] = size(Y); %tamanho do vetor de saidas
12
13 ymed = sum(Y)/tamanho;
14
15 SQe = 0;
16 for k = 1:tamanho
17     SQe = SQe + (Y(k) - Yout(k))^2; %soma para calcular a somatorio dos
18     %quadrados das diferenças de y e yout
19 end
20 %num vale a soma dos quadrados das diferenças entre y e yout
21
22 Syy=0;
23 for w = 1:tamanho
24     Syy = Syy + (Y(w) - ymed)^2; %soma para calcular a soma dos quadrados
25     %das diferenças entre y e ymed
26 end
27 %den vale a soma dos quadrados das diferenças entre y e ymed
28
29 coef = 1 - (SQe/(tamanho - grau + 1))/(Syy/(tamanho - 1)); %coef vale o
30 %coeficiente de determinação ajustado
```

Figura 8. Função *Raj* (retorna coeficiente pela métrica R^2_{aj}).

Fonte: Próprio Autor

A função *erroMedio* pode ser vista na figura 9.

```

1      %Autor: Denilson Gomes Vaz da Silva
2      %Graduando em Engenharia da Computação
3      %Inteligência Computacional - Dr. Jarbas Joaci
4      %Função para determinar o erro medio
5
6      function eM = erroMedio(Y,Yout)
7      -   e = abs(Y - Yout);
8      -   eM = mean(e);

```

Figura 9. Função *erroMedio* (retorna o erro médio obtido).

Fonte: Próprio Autor

4. RESULTADOS

4.1 Regressão Polinomial

Na regressão polinomial, a saída obtida é um polinômio do grau fornecido pelo usuário na execução. Para um polinômio de grau 2, obtemos o seguinte resultado.

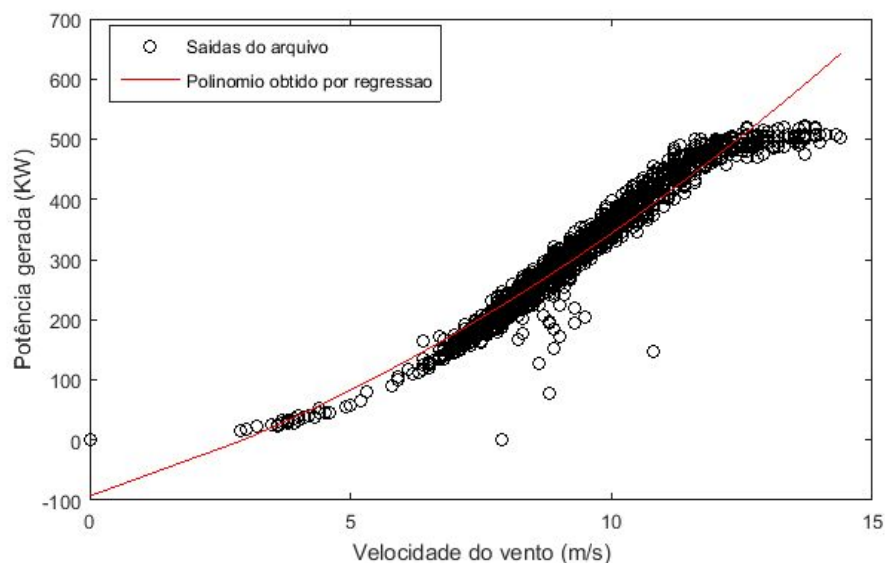


Figura 10. Gráfico para regressão polinomial de grau 2.

Fonte: Próprio Autor

```

Digite o grau do polinômio (entre 2 e 6): 2
Coeficiente de Determinação R^2: 0.94342
Coeficiente de Determinação R^2 Ajustado: 0.9434
Alcançando um erro medio de : 17.1773
fx >> |

```

Figura 11. Coeficientes de determinação e erro médio para polinômio de grau 2. Fonte: Próprio Autor

É interessante notar que o polinômio passou perto da maioria dos pontos da base de dados, porém é possível se aproximar ainda mais. Para a análise da performance de cada polinômio foi construída uma tabela com os coeficientes e o

erro médio para cada polinômio. Estes dados de cada polinômio podem ser vistos na tabela 1.

Grau	Coef. R^2	Coef. R^2 Ajustado	Erro Médio
2	0.94342	0.9434	17.17730
3	0.96902	0.96900	12.07070
4	0.97372	0.97369	10.76100
5	0.97373	0.97368	10.76270
6	0.97375	0.97369	10.80180

Tabela 1. Coeficientes de determinação e erro médio.

Fonte: Próprio Autor

Com esses dados em mãos podemos concluir que o melhor grau para essa regressão polinomial é o grau 4, pois apresenta um erro médio menor do que os outros graus.

O gráfico da regressão polinomial para grau 4 pode ser visto na figura 12.

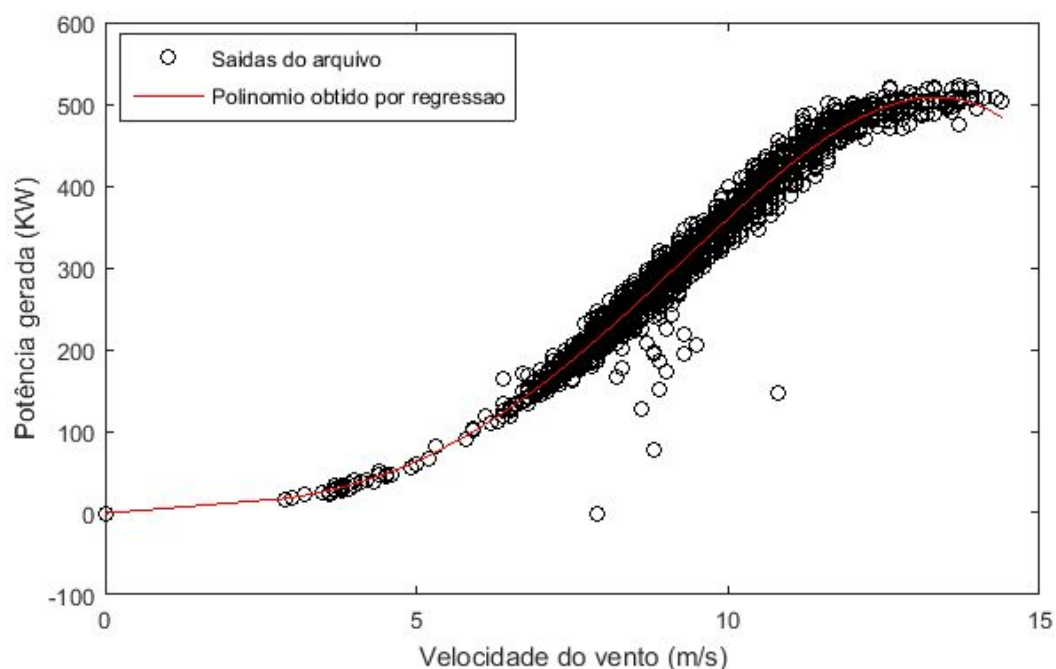


Figura 12. Gráfico para regressão polinomial de grau 4.

Fonte: Próprio Autor

Deve-se ressaltar também que quanto maior o grau do polinômio, maior o poder computacional exigido.

4.2 Regressão Linear

Ao aplicarmos a regressão linear na base D, obtivemos o seguinte resultado:

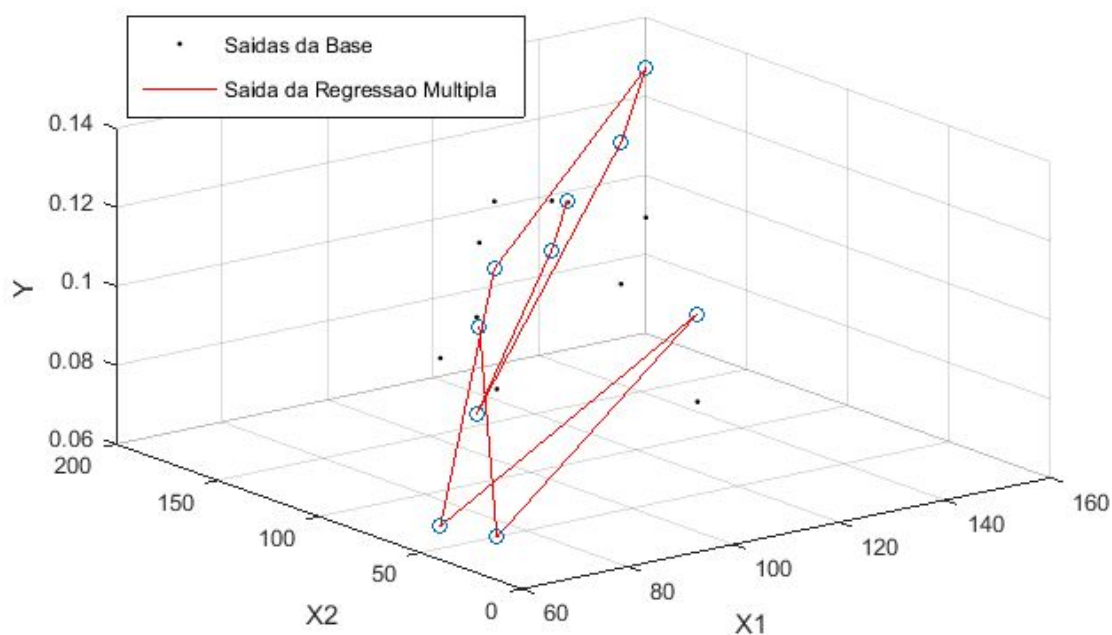


Figura 13. Gráfico para regressão linear na base D.

Fonte: Próprio Autor

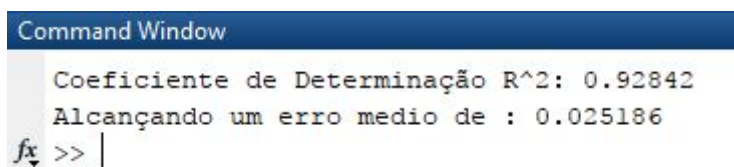


Figura 14. Coeficiente de determinação R^2 e erro médio.

Fonte: Próprio Autor

Mesmo com uma base pequena foi alcançado um coeficiente de determinação razoável. No entanto, o erro médio obtido foi um pouco grande para os valores de saída da base.

REFERÊNCIAS

Mendonça, Melissa.Computação Científica com Matlab. (2013). Disponível em:
<http://mtm.ufsc.br/~melissa/20132/epagri/aula_06.pdf>

Pentenate, Marcelo.Regressão linear e múltipla: entenda as diferenças!. (2013).
Disponível em:
<<http://www.escolaedti.com.br/regressao-linear-e-multipla-entenda-as-diferencas/>>