UNIVERSIDADE FEDERAL DO CEARÁ

CAMPUS SOBRAL

ENGENHARIA DA COMPUTAÇÃO

INTELIGÊNCIA COMPUTACIONAL

DENILSON GOMES VAZ DA SILVA 374872

IMPLEMENTAÇÃO DOS MODELOS DE REGRESSÃO (POLINOMIAL E LINEAR)

SOBRAL

2018

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.

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

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

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

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.bat” 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.

```
%recebe o grau do polinomio a ser gerado
grau = input('Digite o grau do polinômio (entre 2 e 6): ');

load aerogerador.dat %carrega o arquivo

x = aerogerador(:,1); %x recebe a coluna da variavel independente
y = aerogerador(:,2); %y recebe a coluna da variavel dependente

[tamanho,~] = size(x); %recebe o tamanho de entradas

%X é a matriz dos polinomio
X(:,1) = ones(tamanho,1); %para cada entrada atribuimos o valor 1
%na primeira coluna

%Criamos as colunas x, x^2, x^3 ,...,x^grau
for i = 1:grau
    X(:,i+1) = aerogerador(:,1).^(i); %atribuimos a entrada elevada
    %ao devido expoente
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.

```
%operação com as matrizes para encontrarmos as constantes que multiplicam x
a = ((X'*X)^(-1))*X'*y; %a contem as constantes do polinomio

%yout é a saida polinomial obtida por regressão
yout = a(1); %yout recebe o valor da primeira constante

%yout é acrescido das outras componentes
for j = 1:grau
     yout = a(j+1)*x.^(j) + yout; %yout é acrescido da conponente x^j
end
%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.

```
plot(x,y,'ko'); %plota as entradas x e as saidas y do arquivo aerogerador.dat
hold on; %mantem o grafico x vs y
plot(x,yout,'r'); %plota as entradas x e o polinomio obtido por regressao
ylabel('Potência gerada (KW)'); %titulo eixo vertical
xlabel('Velocidade do vento (m/s)'); %titulo eixo horizontal
%legend('Saidas do arquivo', 'Polinomio obtido por regressao', 'tl')

%cálculo de R^2 e R^2ajustado
r1 = R(y,yout); %r1 recebe o coeficiente de determinação
r2 = Raj(y,yout,X); %r2 recebe o coeficiente ajustado
em = erroMedio(y,yout); %em recebe o erro medio

str = ['Coeficiente de Determinação R^2: ' num2str(r1)];
disp(str); %exibe a mensagem acima com o valor de r1

str = ['Coeficiente de Determinação r^2 Ajustado: ' num2str(r2)];
disp(str); %exibe a mensagem acima com o valor de r2
str = ['Erro medio: ' num2str(em)];
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.

```
%base de dados D
D =[122 139 0.115;
 114 126 0.120;
 086 090 0.105;
 134 144 0.090;
 146 163 0.100;
 107 136 0.120;
 068 061 0.105;
 117 062 0.080;
 071 041 0.100;
 098 120 0.115];

x1 = D(:,1); %variavel independente x1
x2 = D(:,2); %variavel independente x2
y = D(:,3); %variavel dependente y

%matriz X
X(:,1) = x1;
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 contêm as constantes que multiplicam as variáveis de entrada. Após obter estas constantes constantes basta multiplicar a matriz X por B e obtemos a saída *yout*. Esse trecho do código é visto na figura 5.

```
B = ((X'*X)^(-1))*X'*y; %B contem as constantes
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.

```
plot3(x1,x2,y,'k.'); %plota as entradas x1 e x2 para as saidas y da base
hold on; %mantem o grafico
grid on; %habilita as linhas de grade
plot3(x1,x2,yout,'r'); %plota as entradas x1 e x2 para a saida obtida por regressao
xlabel('X1'); %eixo x1
ylabel('X2'); %eixo x2
zlabel('Yout'); %eixo yout
%legend('Saidas da Base', 'Saida da Regressao Multipla', 'nl')

%cálculo de R^2
r = R(y,yout); %r recebe o coeficiente de determinação
EM = erroMedio(y,yout); %EM recebe o erro medio

str = ['Coeficiente de Determinação R^2: ' num2str(r)];
disp(str); %exibe a mensagem acima com o valor de r
str = ['Erro Medio: ' num2str(EM)];
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.

```
function coef = R(Y,Yout)
[tamanho,~] = size(Y); %tamanho recebe o numero de saidas

ymed = sum(Y)/tamanho; %ymed recebe o valor medio das saidas Y

SQe = 0;
for k = 1:tamanho
    SQe = SQe + (Y(k) - Yout(k))^2;
end
%SQe recebe o somatorio dos quadrados das diferenças entre y e yout

Syy=0;
for w = 1:tamanho
    Syy = Syy + (Y(w) - ymed)^2;
end
%Syy recebe o somatorio dos quadrados das diferenças entre y e ymed

coef =  1 - SQe/Syy; %coef vale o coeficiente de determinação
```

Figura 7. Função *R* (retorna coeficiente pela métrica R^2). Fonte: Próprio Autor

```
function coef = Raj(Y,Yout,poli)

[~,grau] = size(poli); %grau do polinimio
[tamanho,~] = size(Y); %tamanho do vetor de saidas

ymed = sum(Y)/tamanho;

SQe = 0;
for k = 1:tamanho
    SQe = SQe + (Y(k) - Yout(k))^2; %soma para calcular a somatorio dos
    %quadrados das diferenças de y e yout
end
%num vale a soma dos quadrados das diferenças entre y e yout

Syy=0;
for w = 1:tamanho
    Syy = Syy + (Y(w) - ymed)^2; %soma para calcular a soma dos quadrados
    %das diferenças entre y e ymed
end
%den vale a soma dos quadrados das diferenças entre y e ymed

coef =  1 - (SQe/(tamanho - grau + 1))/(Syy/(tamanho - 1)); %coef vale o
%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.

```
%Autor: Denilson Gomes Vaz da Silva
%Graduando em Engenharia da Computação
%Inteligência Computacional - Dr. Jarbas Joaci
%Função para determinar o erro medio

function eM = erroMedio(Y,Yout)
e = abs(Y - Yout);
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.

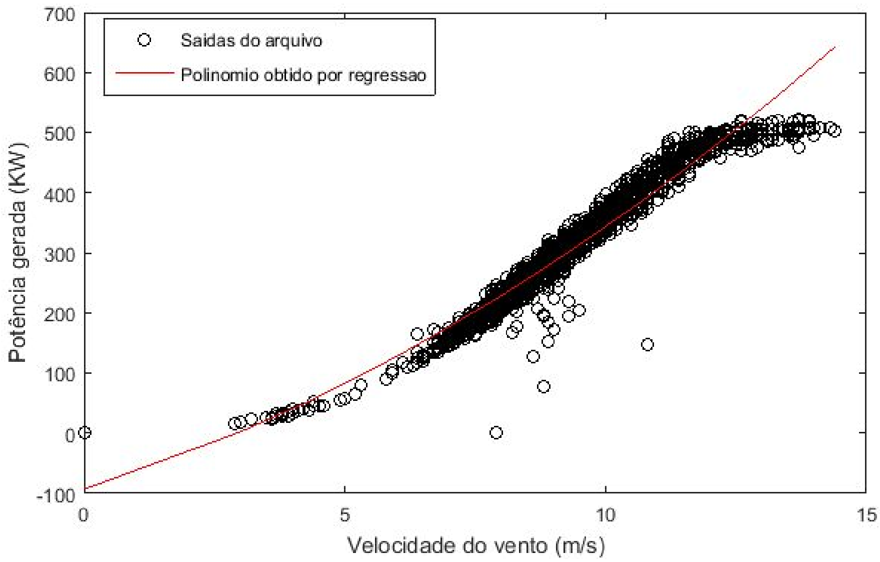

Figura 10. Gráfico para regressão polinomial de grau 2. Fonte: Próprio Autor

```
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
>>
```

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.

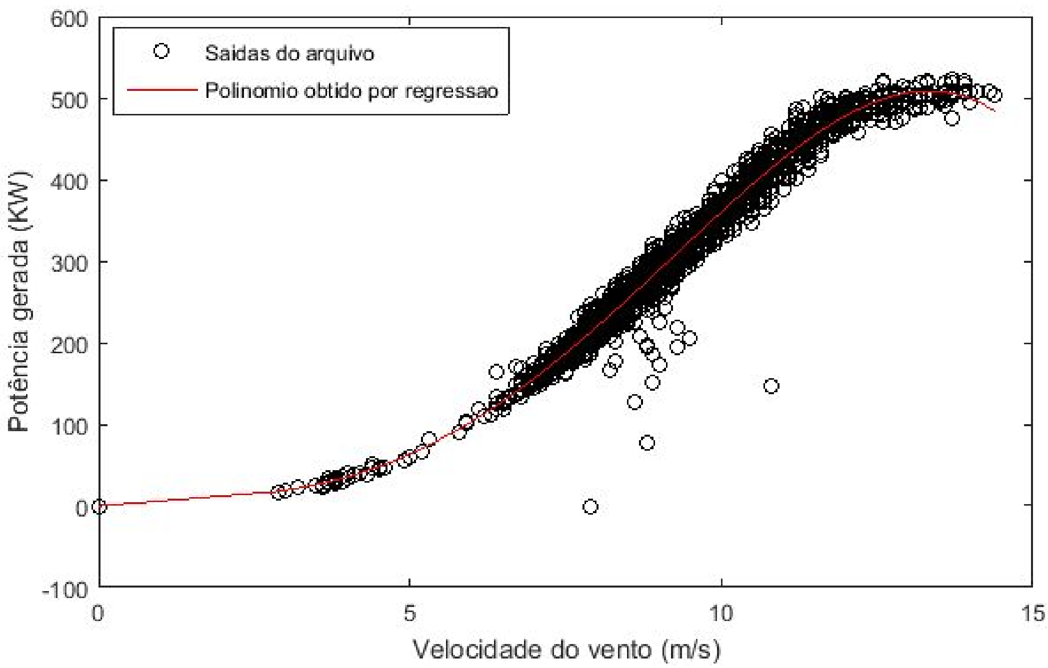

Figura 12. Gráfico para regressão polinomial de grau 4. Fonte: Próprio Autor

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:

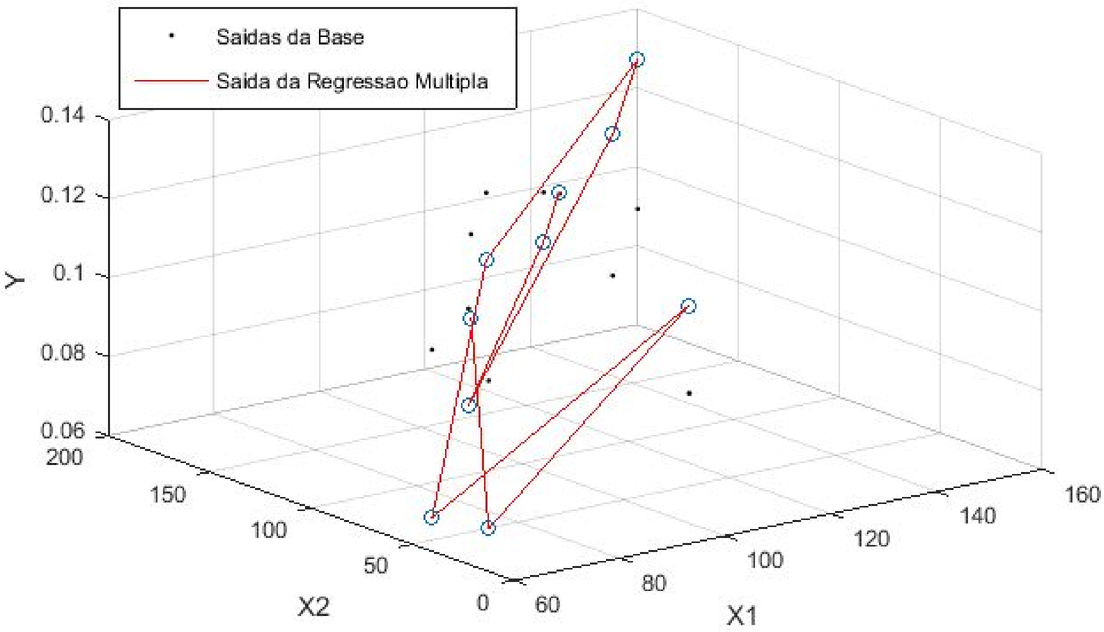

Figura 13. Gráfico para regressão linear na base D. Fonte: Próprio Autor

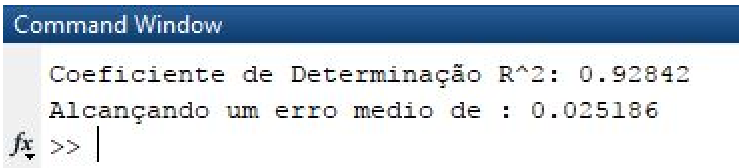

Figura 14. Coeficiente de determinação R^2 e erro médio. Fonte: Próprio Autor

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/>