

Relatório 01

Método dos Mínimos Quadrados – LMS

Regressão Linear

Cristiano Lopes Moreira

Matrícula: 119103-0

Aluno		RA/Matrícula	Professor	Tipo	
Cristiano Lopes Moreira		119103-0	Dr Reinaldo Bianchi	Relatório de implementação	
Data	Versão	Turma	Nome do arquivo		Página
28/10/2019	1	2º. Semestre de 2019	PEL_208_Relatório_01_Cristiano_Moreira.doc		1 (22)

Sumário

1.	Introdução	3
2.	Desenvolvimento teórico	4
2.1.	Regressão Linear Simples	4
2.2.	Regressão Linear Robusta	7
2.3.	Regressão não linear - Quadrática	7
2.4.	Descrição do problema:	8
3.	Proposta de implementação	9
3.1.	Algoritmo de Regressão Linear:	9
3.2.	Pseudocódigos	9
4.	Experimentação e Resultados	11
4.1.	Censo Americano	12
4.2.	Pressão x Temperatura de Ebulição da água nos Alpes	15
4.3.	Livros, palestras e nota	18
5.	Conclusão	22
6.	Referências	22

Aluno		RA/Matrícula		Professor		Tipo	
Cristiano Lopes Moreira		119103-0		Dr Reinaldo Bianchi		Relatório de implementação	
Data	Versão	Turma		Nome do arquivo		Página	
28/10/2019	1	2º. Semestre de 2019		PEL_208_Relatório_01_Cristiano_Moreira.doc		2 (22)	

1. Introdução

Prever, ou predizer, é “ter uma ideia antecipada de (algo que vai acontecer); antever; observar com antecedência e concluir sobre as prováveis consequências; conjecturar, supor, presumir; alcançar o conhecimento de fatos futuros; adivinhar, profetizar, antever; fazer supor; subentender, pressupor; estudar com antecedência; examinar, analisar, avaliar; ver, providenciar com antecedência.” (HOUAISS, 2009). Prefixo ‘pre-’ “de orig. lat., prae-, que ocorre em lat. como preposição de abl., como adv. e como pref. propriamente dito, com a noção de ‘anterioridade, antecipação, adiantamento, diante, superioridade comparativa” (HOUAISS, 2009) e Verbo “ver’ perceber pela visão; enxergar olhar para (alguém, algo ou si próprio); contemplar(-se) distinguir ou alcançar com a vista; avistar, divisar, enxergar; ter conhecimento ou experiência de; submeter à experiência; experimentar” (HOUAISS, 2009).

Uma das principais áreas de estudos da Inteligência Artificial (IA) busca desenvolver modelagens para prever, com o aprendizado estatístico de máquina – SML (Statistical Machine Learning), os comportamentos e fenômenos dos problemas trabalhados através da relação de duas ou mais variáveis, variáveis explicativas, que descrevem um ambiente observado com o objetivo de abastecer os sistemas de tomada de decisão que atuam pela antecipação probabilística, estimativa, do resultado de suas ações. “Um agente estará aprendendo se melhorar o seu desempenho nas tarefas futuras de aprendizagem após fazer observações sobre o mundo.” (RUSSELL; NORVING, 2013)

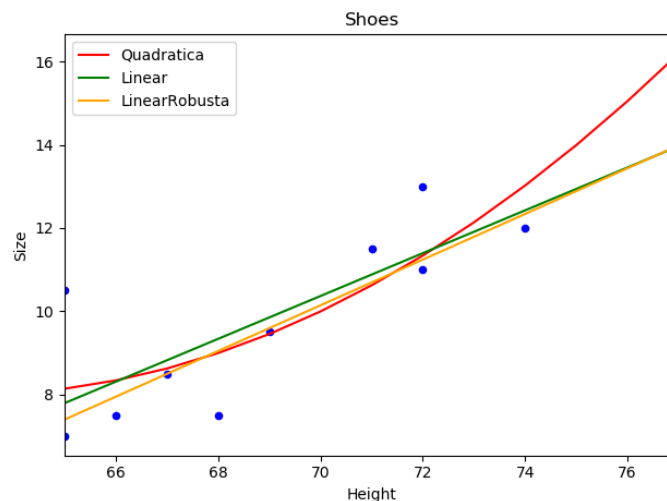
A regressão linear é uma das técnicas matemáticas mais utilizadas na academia para estudar a relação entre duas ou mais variáveis explicativas e uma variável dependente métrica, na qual a variável dependente é o fenômeno em estudo. Este trabalho tem o objetivo de estudar a previsão numérica por regressão linear com o método dos mínimos múltiplos quadrados pelo teorema de Gauss-Markov, em sua forma tradicional paramétrica simples, quadrática e robusta, com resiliência a valores atípicos (outlier).

Aluno		RA/Matrícula	Professor	Tipo	
Cristiano Lopes Moreira		119103-0	Dr Reinaldo Bianchi	Relatório de implementação	
Data	Versão	Turma	Nome do arquivo		Página
28/10/2019	1	2º. Semestre de 2019	PEL_208_Relatório_01_Cristiano_Moreira.doc		3 (22)

2. Desenvolvimento teórico

A regressão linear é um dos algoritmos mais conhecidos e utilizados em estatística e em aprendizado de máquina, sua representação linear é uma equação matemática que melhor descreve uma reta que se encaixa entre os pontos amostrados de uma variável X com saída representando a descrição de um fenômeno.

$$y = f(X) = \alpha + \beta \quad (1)$$



2.1. Regressão Linear Simples

O método dos mínimos quadrados trata de estimar parâmetros α e β pelo quadrado dos resíduos (erros) entre a diferença dos dados observados e seus valores esperados. O método consiste em em problema de regressão, em que a variação em uma variável, chamada variável dependente Y , pode ser parcialmente explicada pela

Aluno		RA/Matrícula	Professor	Tipo	
Cristiano Lopes Moreira		119103-0	Dr Reinaldo Bianchi	Relatório de implementação	
Data	Versão	Turma	Nome do arquivo		Página
28/10/2019	1	2º. Semestre de 2019	PEL_208_Relatório_01_Cristiano_Moreira.doc		4 (22)

variação nas outras variáveis, denominadas explicativas X . Por exemplo, a variação nos resultados do exame Y é causada principalmente pela variação nas leituras e presença em congressos X dos alunos, ou variação na temperatura de ebulição da água Y é devido principalmente a variações na pressão atmosférica X . Dado o valor de X , o a melhor previsão de Y (em termos de erro quadrático médio) é a média $f(X)$ de Y dado X . Dizemos que Y é uma função de X mais ruído/erro, onde a função f é chamada de função regressão calculada a partir da amostragem de i covariáveis e suas respostas $(x_1, y_1), \dots, (x_i, y_i)$:

$$Y = f(X) + \varepsilon \quad (2)$$

Considerando um fenômeno observado em que as variáveis explicativas e variáveis dependentes são mensuradas pontualmente, podemos descrever a representação de forma matricial de X como sendo:

$$X^T = (X_1, X_2, \dots, X_p) \quad (3)$$

Ou

$$\hat{Y} = \hat{\beta}_0 + \sum_{i=1}^p X_i \hat{\beta}_i \quad (4)$$

Suponha que Y seja conhecido até um número finito de pontos p dos parâmetros $\beta = (\beta_1, \dots, \beta_p)$, ou seja, $Y = X\beta$. Estimamos β pelo valor de coeficiente $\hat{\beta}$ que melhor se ajusta aos dados. O coeficiente de estimativa calculado pelo método dos mínimos quadrados, denotado por $\hat{\beta}$, é o valor de b que minimiza os resíduos (r) na somatória abaixo considerando todas as possibilidades de b

Aluno		RA/Matrícula		Professor		Tipo	
Cristiano Lopes Moreira		119103-0		Dr Reinaldo Bianchi		Relatório de implementação	
Data	Versão	Turma		Nome do arquivo			Página
28/10/2019	1	2º. Semestre de 2019		PEL_208_Relatório_01_Cristiano_Moreira.doc			5 (22)

$$\text{resíduo}(f(x)) = \sum_{i=1}^p (y_i - f(x_i))^2 \quad (5)$$

Sendo: $f(x) = a + bx \quad (6)$

Critérios dos Mínimos Quadrados $MMQ = \min \sum_{i=1}^p (y_i - f(x_i))^2 \quad (7)$

Em notação matricial podemos definir a equação 8 levando a equação 7 à 4 e, pela diferenciação em relação ao coeficiente angular β , que nos leva ao ponto de mínimo da função quando seu resultado é 0, obter a equação 10 com a relação dos coeficientes angulares da regressão linear

$$\text{resíduo}(f(x)) = (y - X\beta)^T (y - X\beta) \quad (8)$$

$$\frac{\partial r}{\partial \beta} = -2X^T (y - X\beta) \quad (9)$$

Coeficientes angulares $\hat{\beta} = (X^T X)^{-1} X^T y \quad (10)$

A busca pelo critério dos mínimos quadrados é uma medida de ajuste computacionalmente conveniente. Corresponde à estimativa de probabilidade máxima quando o resíduo é normalmente distribuído com variações iguais. Em alguns casos outras medidas de ajuste são usadas, por exemplo, desvios menos absolutos, mais robustos em relação aos valores extremos, de forma a minimizar os impactos com valores atípicos (outlier), nomeada Regressão Linear Robusta.

Aluno		RA/Matrícula	Professor	Tipo	
Cristiano Lopes Moreira		119103-0	Dr Reinaldo Bianchi	Relatório de implementação	
Data	Versão	Turma	Nome do arquivo		Página
28/10/2019	1	2º. Semestre de 2019	PEL_208_Relatório_01_Cristiano_Moreira.doc		6 (22)

2.2. Regressão Linear Robusta

Em geral os resíduos/erros de uma amostra seguem uma distribuição normal e sobre uma variância previsível, porém ocorrem valores discrepantes que surgem devido a mudanças no comportamento do sistema, erro humano, erro de instrumento ou simplesmente devido a desvios naturais. Para minimizar a influência de valores discrepantes, é possível ajustar os dados usando métodos de regressão robustos.

O método robusto dos pesos bis-quadrados: minimiza uma soma ponderada de quadrados, onde um peso é atribuído a cada ponto de acordo com quão longe o ponto está da linha ajustada.

Este método reduz as interferências de outliers, ajustando a regressão com pesos (W) sem que seja necessário remover o ponto discrepante.

Método robusto
$$\hat{\beta} = (X^T W X)^{-1} X^T W y \quad (11)$$

2.3. Regressão não linear - Quadrática

Nas diversas áreas do conhecimento seria estranho supor que determinada variável dependente em estudo seja influenciada apenas por meio da relação linear de variáveis explicativas. Muitas relações comportam-se por meio de fórmulas funcionais não lineares e são de relevante importância para a compreensão do comportamento dos mais diversos fenômenos.

Assim como na forma linear o coeficiente β indica o efeito marginal da variação de Y em relação a variação de X . No modelo de regressão polinomial Quadrática o β representa o efeito marginal da variação de Y do quadrado da variável explicativa, ou

Aluno Cristiano Lopes Moreira		RA/Matrícula 119103-0	Professor Dr Reinaldo Bianchi	Tipo Relatório de implementação	
Data 28/10/2019	Versão 1	Turma 2º. Semestre de 2019	Nome do arquivo PEL_208_Relatório_01_Cristiano_Moreira.doc		Página 7 (22)

ainda no exponencial, o coeficiente β da regressão pode ser interpretado como uma taxa de crescimento.

“A definição da melhor forma funcional é, por vezes, uma questão empírica a ser decidida a favor do melhor ajuste das variáveis que, por vezes, passam por transformações” (FÁVERO et al., 2009)

Algumas das formas funcionais mais utilizadas:

Forma Funcional	Modelo
Linear	$Y = \alpha + \beta X$
Exponencial	$\ln Y = \ln \alpha + \beta X$
Logarítma	$\ln Y = \ln \alpha + \beta \ln X$
Semilogarítma	$\ln Y = \alpha + \beta X$
Inversa	$Y = \alpha + \beta (1/X)$
Quadrática	$Y = \alpha + \beta X^2$
Cúbica	$Y = \alpha + \beta X^3$
Logística	$Y = \frac{1}{1 + e^{-(\alpha + \beta X)}}$

Fonte: (FÁVERO et al., 2009)

2.4. Descrição do problema:

Deseja-se conhecer e comparar os sistemas de regressão linear, simples, quadrática e robusta, para as bases de dados de crescimento demográfico dos Estados Unidos da América, da relação medida entre a temperatura de ebulição da água nos Alpes, e da relação de notas dos alunos de uma determinada instituição que comparou a nota pela frequência do aluno em seminários e os livros esse aluno pegou na biblioteca; deforma a verificar o melhor modelo de regressão para cada uma destas bases de dados.

Aluno		RA/Matrícula	Professor	Tipo	
Cristiano Lopes Moreira		119103-0	Dr Reinaldo Bianchi	Relatório de implementação	
Data	Versão	Turma	Nome do arquivo		Página
28/10/2019	1	2º. Semestre de 2019	PEL_208_Relatório_01_Cristiano_Moreira.doc		8 (22)

3. Proposta de implementação

3.1. Algoritmo de Regressão Linear:

O algoritmo para realizar o cálculo dos coeficientes de uma regressão linear irá utilizar, em sua base, as regras e operações matemáticas de matrizes, para tal será desenvolvido as rotinas: `multiplica_matriz`, rotina para multiplicação de 2 matrizes, recebe 2 Matrizes ($n \times m$) e ($m \times p$) e retorna uma matriz ($n \times p$); `matriz_transposta`, rotina para gerar a Matriz transposta, recebe uma matriz A ($n \times m$) e retorna uma matriz ($m \times n$); `matriz_determinante`, rotina para calcular o determinante de uma matriz, recebe uma matriz A e retorna seu determinante; `matriz_cofator`, rotina para calcular o complemento algébrico (cofator) de uma matriz, recebe uma matriz A e uma sua posição ($n \times m$) e retorna o valor numérico do cofator na posição; `matriz_Adj`, rotina para formação da matriz de cofatores (matriz Adjacente), recebe uma matriz A e retorna uma matriz de cofatores; `matriz_Inversa`, Rotina para formação da matriz inversa, recebe uma matriz A e retorna a matriz inversa A^{-1} .

3.2. Pseudocódigos

```

multiplica_matriz(matrizA, matrizB)
    recebe matrizes  $A_{(n, m)}$  e  $B_{(m, n)}$ 
    verifica se colunas de  $A$  ( $m$ ) = Linhas de  $B$  ( $m$ )
    loop para cada item da matriz  $A_{(n, m)}$ 
        matriz resultante  $C_{(i, j)} \leftarrow A_{(i, j)} \times B_{(j, i)}$ 
    retorna  $C$ 

```

```

matriz_transposta(matrizA)
    recebe matrizes  $A_{(n, m)}$ 
    loop para cada item da matriz  $A_{(n, m)}$ 
        matriz resultante  $C_{(j, i)} \leftarrow A_{(i, j)}$ 
    retorna  $C$ 

```

Aluno		RA/Matrícula	Professor	Tipo	
Cristiano Lopes Moreira		119103-0	Dr Reinaldo Bianchi	Relatório de implementação	
Data	Versão	Turma	Nome do arquivo		Página
28/10/2019	1	2º. Semestre de 2019	PEL_208_Relatório_01_Cristiano_Moreira.doc		9 (22)

```
matriz_determinante(matrizA)
  recebe matrizes  $A_{(n, m)}$ 
  Se Linhas de A (n) e Colunas (m) =1
    Determinante  $\leftarrow A_{(0, 0)}$ 
  Se Linhas de A (n) e Colunas (m) =2
    Determinante  $\leftarrow A_{(0, 0)} \times A_{(1, 1)} - A_{(0, 1)} \times A_{(1, 0)}$ 
  Se Linhas de A (n) e Colunas (m) >2
    Determinante  $\leftarrow$  determinante por teorema de Laplace
  retorna Determinante
```

```
matriz_cofator(matrizA, índice_i, índice_j)
  MatrizC  $\leftarrow$  matrizA excluindo a linha i e coluna j
  DeterminanteC  $\leftarrow$  matriz_determinante(MatrizC)
  Cofator  $\leftarrow ((-1)^{(i+1)+(j+1)}) \times$  DeterminanteC
  Retorna Cofator
```

```
matriz_Adjacente(matrizA)
  recebe matrizes  $A_{(n, m)}$ 
  loop para cada item da matriz  $A_{(n, m)}$ 
    matriz resultante  $C_{(i, j)} \leftarrow$  matriz_cofator ( $matrizA_{(i, j)}$ )
  retorna matriz_transposta( matriz resultante C )
```

```
matriz_inversa(matrizA)
  recebe matriz  $A_{(n, m)}$ 
  matrizAdj  $\leftarrow$  matriz_adjacente( $A_{(n, m)}$ )
  DeterminanteA  $\leftarrow$  matriz_determinante( $A_{(n, m)}$ )
  loop para cada item da matriz  $A_{(n, m)}$ 
    matriz inversa  $C_{(i, j)} \leftarrow$  matrizAdj $_{(i, j)} / (1/\text{DeterminanteA})$ 
  retorna matriz inversa C
```

```
regressao(base de dados)
  recebe base de dados
  X  $\leftarrow$  variáveis explicativas da base de dados
  Y  $\leftarrow$  variável dependente da base de dados
   $\hat{\beta} = (X^T X)^{-1} X^T y$ 
  retorna  $\hat{\beta}$ 
```

Aluno		RA/Matrícula	Professor	Tipo	
Cristiano Lopes Moreira		119103-0	Dr Reinaldo Bianchi	Relatório de implementação	
Data	Versão	Turma	Nome do arquivo		Página
28/10/2019	1	2º. Semestre de 2019	PEL_208_Relatório_01_Cristiano_Moreira.doc		10 (22)

4. Experimentação e Resultados

Para realizar a regressão pelos 3 métodos, linear simples, robusta e quadrática, foi realizada a implementação em Python confrontando os diferentes resultados com três bases de dados:

- Dados do Censo Americano, com o tamanho da população dos Estados Unidos entre 1900 e 2000;
- Base de dados com 17 medições de temperatura x pressão nos Alpes;
- Base de dados com 23 amostras do resultado dos alunos sendo de 2 variáveis explicativas (livros e palestras) e 1 relacionada (nota).

Ambiente:

PyCharm 2019.2.2 (Professional Edition) Build#PY-192.6603.34

Python 3.7.5 (tags/v3.7.5:5c02a39a0b, Oct 15 2019, 01:31:54) on win32

Bibliotecas:

matplotlib-3.1.1	(utilizado para plotagem de gráficos)
pandas-0.25.2	(suporte à plotagem de gráficos)
xlrd-1.2.0	(leitura de arquivos do Excel - base de dados)

Base de Dados:

bianchiDbAula1.xlsx	(Base uni-variáveis)
bianchiDbAula1MultiVariada.xlsx	(base multivariáveis)

Aluno		RA/Matrícula	Professor	Tipo	
Cristiano Lopes Moreira		119103-0	Dr Reinaldo Bianchi	Relatório de implementação	
Data	Versão	Turma	Nome do arquivo		Página
28/10/2019	1	2º. Semestre de 2019	PEL_208_Relatório_01_Cristiano_Moreira.doc		11 (22)

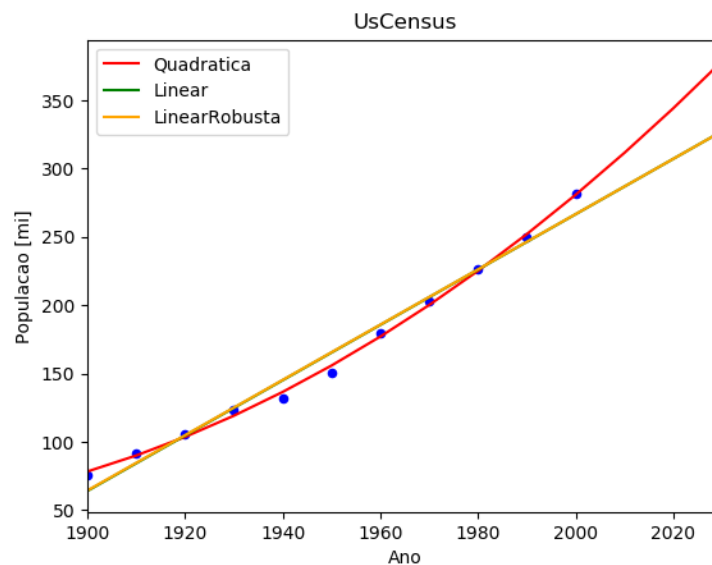
4.1. Censo Americano

Base de dados:

Censo dos EUA – Valores Mensurados	
Ano	População [Mil]
1900	76.00
1910	91.97
1920	105.71
1930	123.20
1940	131.67
1950	150.70
1960	179.32
1970	203.21
1980	226.51
1990	249.63
2000	281.42

Aluno		RA/Matrícula	Professor	Tipo	
Cristiano Lopes Moreira		119103-0	Dr Reinaldo Bianchi	Relatório de implementação	
Data	Versão	Turma	Nome do arquivo		Página
28/10/2019	1	2º. Semestre de 2019	PEL_208_Relatório_01_Cristiano_Moreira.doc		12 (22)

Resultados:



A regressão dos valores demográficos dos Estados Unidos da América mostra-se mais ajustada para o modelo de regressão quadrática que contempla a maior parte dos pontos dentro da projeção da curva. Os modelos lineares, simples e robusto, são bem semelhantes e não demonstrarão as variações posteriores ao ano de 1980 e não se aproximaram do resultado real no ano de 2010.

Coeficientes - $\hat{\beta}$ Censo dos EUA		
Regressão Linear		
Intercepto	$\hat{\beta}_0$	-81.064
Coeficiente angular de Crescimento da População	$\hat{\beta}_1$	0.523
Regressão Linear Robusta		
Intercepto	$\hat{\beta}_0$	-81.367
Coeficiente angular de Crescimento da População	$\hat{\beta}_1$	0.523
Regressão Quadrática		
Intercepto	$\hat{\beta}_0$	38.829
Coeficiente angular de Crescimento da População	$\hat{\beta}_1$	-0.6548
Coeficiente angular de Crescimento Quadrático	$\hat{\beta}_2$	0.002889

Aluno		RA/Matrícula	Professor	Tipo	
Cristiano Lopes Moreira		119103-0	Dr Reinaldo Bianchi	Relatório de implementação	
Data	Versão	Turma	Nome do arquivo		Página
28/10/2019	1	2º. Semestre de 2019	PEL_208_Relatório_01_Cristiano_Moreira.doc		13 (22)

Censo dos EUA – População [Mil] Modelos de Regressão			
Ano	Linear Simples	Linear Robusta	Quadrática
1900	64.13	64.46	78.36
1910	84.38	84.69	90.08
1920	104.64	104.92	103.68
1930	124.89	125.14	119.19
1940	145.14	145.37	136.60
1950	165.39	165.60	155.90
1960	185.65	185.83	177.11
1970	205.90	206.05	200.21
1980	226.15	226.28	225.20
1990	246.41	246.51	252.10
2000	266.66	266.74	280.90
2010	286.91	286.97	311.59
2020	307.17	307.19	344.18
2030	327.42	327.42	378.67

Em 2010 o Censo americano reportou a população com 308,54 milhões de habitantes, mostrando que o modelo quadrático obteve um erro 0.98% e os modelos lineares um erro 6.92%, 7 vezes maior que o modelo quadrático.

Aluno		RA/Matrícula		Professor		Tipo	
Cristiano Lopes Moreira		119103-0		Dr Reinaldo Bianchi		Relatório de implementação	
Data	Versão	Turma		Nome do arquivo		Página	
28/10/2019	1	2º. Semestre de 2019		PEL_208_Relatório_01_Cristiano_Moreira.doc		14 (22)	

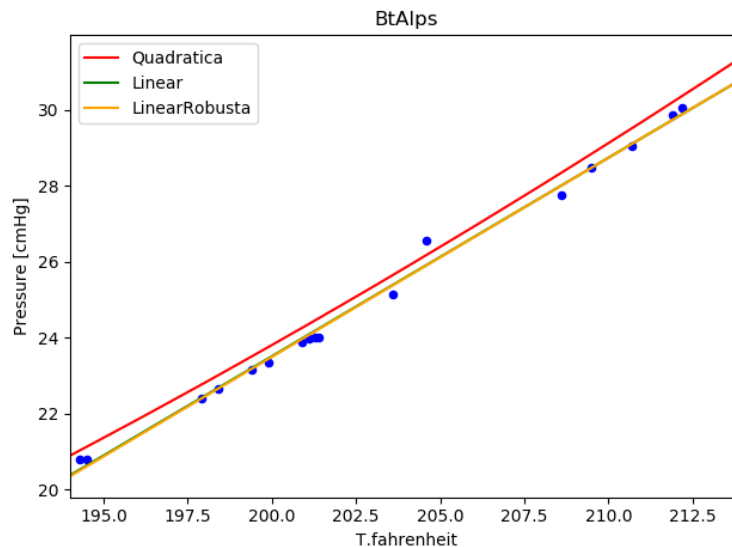
4.2. Pressão x Temperatura de Ebulição da água nos Alpes

Base de dados:

Ebulição da água – Valores Mensurados	
Temperatura [Fº]	Pressão [cmHg]
194.5	20.79
194.3	20.79
197.9	22.4
198.4	22.67
199.4	23.15
199.9	23.35
200.9	23.89
201.1	23.99
201.4	24.02
201.3	24.01
203.6	25.14
204.6	26.57
209.5	28.49
208.6	27.76
210.7	29.04
211.9	29.88
212.2	30.06

Aluno		RA/Matrícula	Professor	Tipo	
Cristiano Lopes Moreira		119103-0	Dr Reinaldo Bianchi	Relatório de implementação	
Data	Versão	Turma	Nome do arquivo		Página
28/10/2019	1	2º. Semestre de 2019	PEL_208_Relatório_01_Cristiano_Moreira.doc		15 (22)

Resultados:



A regressão dos valores de pressão atmosférica x temperatura de ebulição da água nos Alpes mostra-se mais ajustada para o modelo de regressão linear robusta que contempla a maior parte dos pontos dentro da projeção da curva, atenuando a possível distorção apresentada pelo ponto à temperatura de 205 F°.

Coeficientes - $\hat{\beta}$ Ebulição da água			
Regressão Linear			
Intercepto	$\hat{\beta}_0$		-81.064
Coeficiente angular de T[F°]	$\hat{\beta}_1$		0.523
Regressão Linear Robusta			
Intercepto	$\hat{\beta}_0$		-81.367
Coeficiente angular de T[F°]	$\hat{\beta}_1$		0.523
Regressão Quadrática			
Intercepto	$\hat{\beta}_0$		38.829
Coeficiente angular de T[F°]	$\hat{\beta}_1$		-0.6548
Coeficiente angular de T[F°] ²	$\hat{\beta}_2$		0.002889

Aluno		RA/Matrícula	Professor	Tipo	
Cristiano Lopes Moreira		119103-0	Dr Reinaldo Bianchi	Relatório de implementação	
Data	Versão	Turma	Nome do arquivo		Página
28/10/2019	1	2º. Semestre de 2019	PEL_208_Relatório_01_Cristiano_Moreira.doc		16 (22)

Ebulição da água – Pressão [cmHg] Modelos de Regressão			
Temperatura [F°]	Linear Simples	Linear Robusta	Quadrática
194	20.90	20.38	20.35
195	21.37	20.90	20.88
196	21.84	21.42	21.40
197	22.33	21.95	21.93
198	22.81	22.47	22.45
199	23.31	22.99	22.97
200	23.81	23.51	23.50
201	24.31	24.04	24.02
202	24.83	24.56	24.55
203	25.34	25.08	25.07
204	25.87	25.61	25.60
205	26.40	26.13	26.12
206	26.93	26.65	26.64
207	27.47	27.18	27.17
208	28.02	27.70	27.69
209	28.57	28.22	28.22
210	29.13	28.74	28.74
211	29.69	29.27	29.27
212	30.26	29.79	29.79
213	30.83	30.31	30.31

O modelo apresenta um intercepto negativo, -81.367, o que indica uma pressão negativa para a ebulição da água à temperatura de 0 F° sendo uma possível distorção no modelo, que como sabemos pela física deveria ser um modelo exponencial quando falamos da relação entre a pressão e a temperatura de ebulição da água.

Aluno		RA/Matrícula		Professor		Tipo	
Cristiano Lopes Moreira		119103-0		Dr Reinaldo Bianchi		Relatório de implementação	
Data	Versão	Turma		Nome do arquivo		Página	
28/10/2019	1	2º. Semestre de 2019		PEL_208_Relatório_01_Cristiano_Moreira.doc		17 (22)	

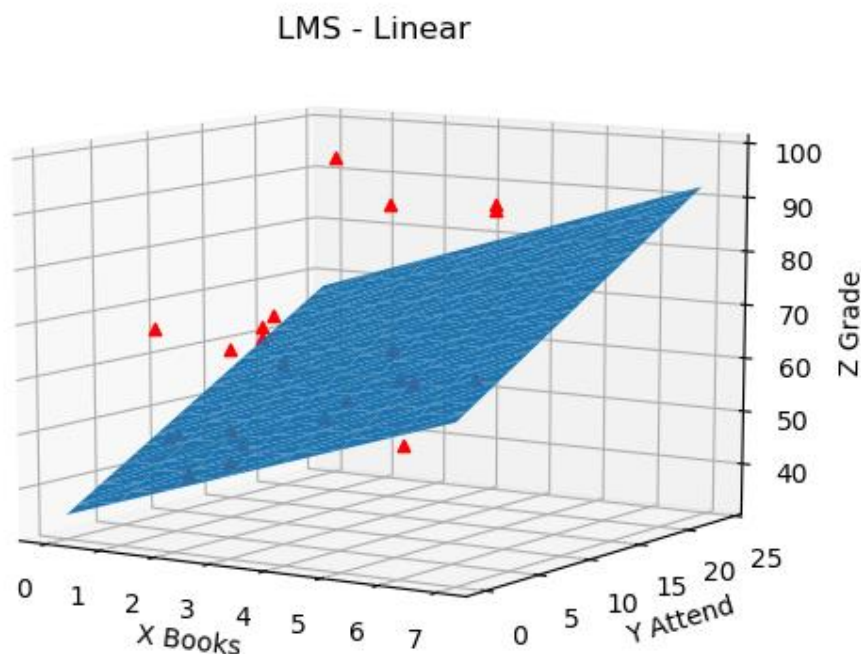
4.3. Livros, palestras e nota

Base de dados:

Livros x Palestras x Notas		
Nota	Livros	Palestras
45	0	9
57	1	15
45	0	10
51	2	16
65	4	10
88	4	20
44	1	11
87	4	20
89	3	15
59	0	15
66	2	8
65	1	13
56	4	18
47	1	10
66	0	8
41	1	10
56	3	16
37	0	11
45	1	19
58	4	12
47	4	11
64	0	19
97	2	15

Aluno		RA/Matrícula		Professor		Tipo	
Cristiano Lopes Moreira		119103-0		Dr Reinaldo Bianchi		Relatório de implementação	
Data	Versão	Turma		Nome do arquivo		Página	
28/10/2019	1	2º. Semestre de 2019		PEL_208_Relatório_01_Cristiano_Moreira.doc		18 (22)	

Resultados:

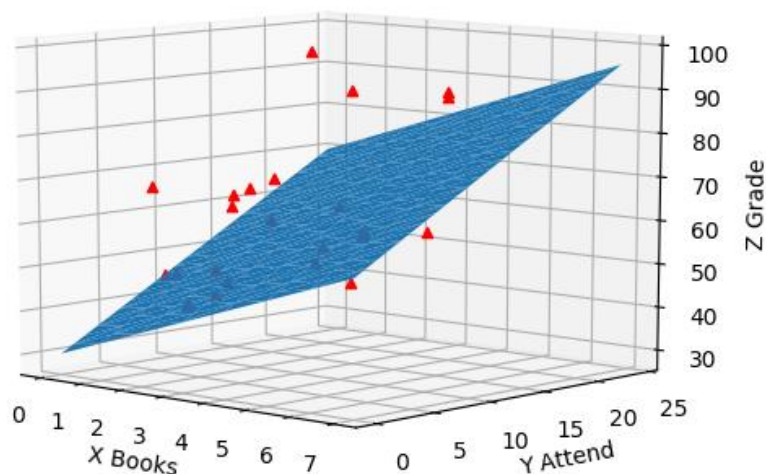


A regressão dos linear das evidências de leitura e participação de palestras pela nota do aluno mostra que, pelo coeficiente angular da variável explicativa livros, a leitura traz 2,5 vezes mais resultados nas notas dos alunos que a participação em palestras. O modelo é aparentemente distorcido pelos outliers de nota 89 e 97, que com poucas leituras obtiveram grande resultado em notas, dentre outros com nenhuma leitura.

Coeficientes - $\hat{\beta}$ Eficiência de alunos			
Regressão Linear			
Intercepto	$\hat{\beta}_0$		34.942
Coeficiente angular de Livros	$\hat{\beta}_1$		3.501
Coeficiente angular de Palestras	$\hat{\beta}_2$		1.364

Aluno		RA/Matrícula		Professor		Tipo	
Cristiano Lopes Moreira		119103-0		Dr Reinaldo Bianchi		Relatório de implementação	
Data	Versão	Turma		Nome do arquivo			Página
28/10/2019	1	2º. Semestre de 2019		PEL_208_Relatório_01_Cristiano_Moreira.doc			19 (22)

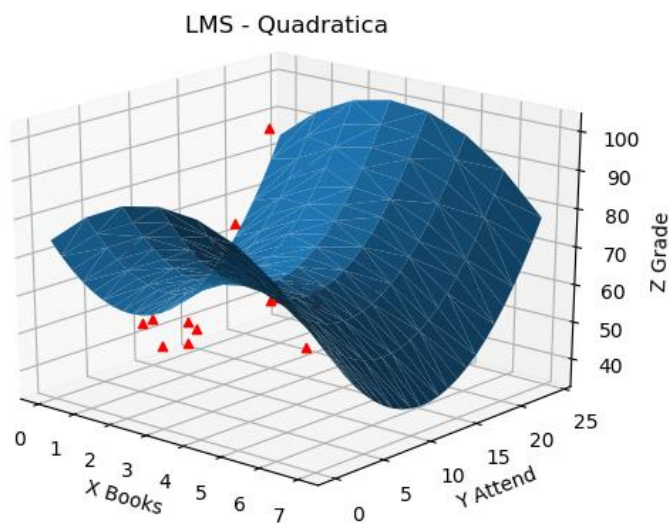
LMS - Robusto



A regressão dos linear robusta das evidências de leitura e participação de palestras pela nota do aluno mostra a grande eficiência da leitura com 2,25 vezes mais resultados nas notas dos alunos que a participação em palestras, assim como no modelo de regressão linear simples, e ponderada para os casos de outliers com poucos livros lidos.

Coeficientes - $\hat{\beta}$ Eficiência de alunos			
Regressão Linear Robusta			
Intercepto	$\hat{\beta}_0$		30.17
Coeficiente angular de Livros	$\hat{\beta}_1$		3.707
Coeficiente angular de Palestras	$\hat{\beta}_2$		1.646

Aluno		RA/Matrícula		Professor		Tipo	
Cristiano Lopes Moreira		119103-0		Dr Reinaldo Bianchi		Relatório de implementação	
Data	Versão	Turma		Nome do arquivo			Página
28/10/2019	1	2º. Semestre de 2019		PEL_208_Relatório_01_Cristiano_Moreira.doc			20 (22)



A regressão Quadrática das evidências de leitura e participação de palestras pela nota do aluno trás um coeficiente angular para a variável explicativa livro positivo, porém mostra também um coeficiente angular negativo para as palestras, o que não aparenta ser real para um modelo de educação.

Este modelo aparenta contemplar a maior parte dos pontos próximos à superfície do gráfico, porém é necessário mais amostras e um conhecimento mais profundo da base de dados para tomar alguma conclusão do motivo dos coeficientes negativos para a variável explicativa presença em palestras.

Coeficientes - $\hat{\beta}$ Eficiência de alunos			
Regressão Quadrática			
Intercepto	$\hat{\beta}_0$		74.468
Coeficiente angular de Livros	$\hat{\beta}_1$		9.773
Coeficiente angular de Palestras	$\hat{\beta}_2$		-5.136
Coeficiente angular de Livros ²	$\hat{\beta}_3$		-1.531
Coeficiente angular de Palestras ²	$\hat{\beta}_4$		0.231

Aluno		RA/Matrícula		Professor		Tipo	
Cristiano Lopes Moreira		119103-0		Dr Reinaldo Bianchi		Relatório de implementação	
Data	Versão	Turma		Nome do arquivo		Página	
28/10/2019	1	2º. Semestre de 2019		PEL_208_Relatório_01_Cristiano_Moreira.doc		21 (22)	

5. Conclusão

O modelo de regressão linear por mínimos quadrados é uma das técnicas mais utilizadas por sua praticidade, eficiência e resultados consistentes.

Essa técnica oferece uma oportunidade para a criação de modelos estatísticos que explicitam a influência relativa de cada parâmetro (variável explicativa) sobre o fenômeno dependente (variável dependente) e possibilita a elaboração de previsões e de simulações do comportamento de um ambiente em função de possíveis ações, valores de uma variável explicativa, que um sistema de tomada de decisão pode ter.

Concluindo, a técnica de regressão utilizada, mesmo para mecanismos de aprendizado estatístico de máquina, deve sempre ter, pelo pesquisador/desenvolvedor, uma análise crítica ao se incluir variáveis explicativas e definir a forma funcional da regressão para evitar a criação de um modelo inserto e consequente tomada de decisões baseadas em previsões incorretas.

6. Referências

- [1] HOUAISS, A.; VILLAR, M. S. **Dicionário Houaiss da Língua Portuguesa**. Rio de Janeiro: Objetiva, 2009 Disponível em: <<https://houaiss.uol.com.br>> Acesso em 29 de Outubro de 2019.
- [2] RUSSELL, Stuart; NORVING, Peter. **Inteligência Artificial**: Tradução da Terceira Edição. 3. ed. Rio de Janeiro: Elsevier, 2013.
- [3] FÁVERO, Luiz Paulo et al. **Análise de dados**: modelagem multivariada para tomada de decisões. Rio de Janeiro: Elsevier, 2009.

Aluno		RA/Matrícula	Professor	Tipo	
Cristiano Lopes Moreira		119103-0	Dr Reinaldo Bianchi	Relatório de implementação	
Data	Versão	Turma	Nome do arquivo		Página
28/10/2019	1	2º. Semestre de 2019	PEL_208_Relatório_01_Cristiano_Moreira.doc		22 (22)