Aluno: Felippe Trigueiros Angelo - 210479

Parte 1 - Atividades Teóricas

01

x | y	y = 0	y = 1
x = 0	1/6	3/8
x = 1	1/8	1/3

a)

$$P(x) = \begin{cases} 1/6 + 3/8, & x = 0 \\ 1/8 + 1/3, & x = 1 \end{cases} = \begin{cases} 13/24, & x = 0 \\ 11/24, & x = 1 \end{cases}$$

$$P(y) = \begin{cases} 1/6 + 1/8, & x = 0 \\ 3/8 + 1/3, & x = 1 \end{cases} = \begin{cases} 7/24, & y = 0 \\ 17/24, & y = 1 \end{cases}$$

b)

$$P(x=0 \mid y=0) = \frac{P(x=0 \cap y=0)}{P(y=0)} = \frac{1/6}{7/24} = \frac{4}{7}$$

c)

$$E(x) = 0 \cdot \frac{13}{24} + \frac{11}{24} = \frac{11}{24}$$

$$E(y) = 0 \cdot \frac{7}{24} + \frac{17}{24} = \frac{17}{24}$$

d)

$$P(x=0)\,P(y=0) = \frac{13}{24} \cdot \frac{7}{24} = \frac{91}{576} \neq P(x=0, y=0)$$

$$P(x=1)\,P(y=0) = \frac{11}{24} \cdot \frac{7}{24} = \frac{77}{576} \neq P(x=1, y=0)$$

$P(X=0)\,P(Y=0) = \frac{13}{24} \cdot \frac{17}{24} = \frac{221}{576} \neq P(X=0, Y=0)$

$P(X=1)\,P(Y=0) = \frac{11}{24} \cdot \frac{17}{24} = \frac{187}{576} \neq P(X=1, Y=0)$

Não são [illegible] tipo da [illegible]!!

02.

X | Y	Y = 0	Y = 1
X = 0	0	1/4
X = 1	3/8	3/8

a) $H(X) = -\left[P(X=0)\log P(X=0) + P(X=1)\log P(X=1)\right]$

$= -\left[\frac{1}{4}\log\frac{1}{4} + \frac{6}{8}\log\frac{6}{8}\right] = 0{,}8113 \text{ bits}$

$H(Y) = -\left[P(Y=0)\log P(Y=0) + P(Y=1)\log P(Y=1)\right]$

$= -\left[\frac{3}{8}\log\frac{3}{8} + \frac{5}{8}\log\frac{5}{8}\right] = 0{,}7916 \text{ bits}$

$H(X,Y) = -\left[\frac{1}{4}\log\frac{1}{4} + \frac{3}{8}\log\frac{3}{8} + \frac{3}{8}\log\frac{3}{8}\right] = 1{,}5613 \text{ bits}$

b) $H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$

$H(X|Y) = H(X,Y) - H(Y) = 1{,}5613 - 0{,}7916 = 0{,}7697 \text{ bits}$

$H(Y|X) = H(X,Y) - H(X) = 1{,}5613 - 0{,}8113 = 0{,}75 \text{ bits}$

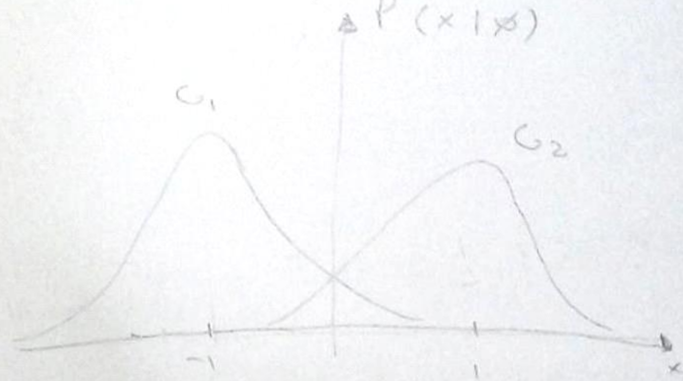
c) $I(X,Y) = H(X) - H(X|Y) = 0{,}8113 - 0{,}7697 = 0{,}0416 \text{ bits}$

03. Classificação 2 classes C_1 e C_2 com um único atributo.

a)

$$C_1 \sim \frac{1}{\sqrt{2\pi}} \exp\left[\frac{-(x+1)^2}{2}\right]$$

$$C_2 \sim \frac{1}{\sqrt{2\pi}} \exp\left[\frac{-(x-1)^2}{2}\right]$$



Precisamos estimar o valor do limiar x. Para tal, será utilizado o estimador por máxima verossimilhança.

Para o caso em questão, o estimador $L(x, \phi)$ será máximo quando x estiver na classe ϕ. Assim:

$L(x, C_1) > L(x, C_2) \rightarrow x : C_1$

$L(x, C_1) < L(x, C_2) \rightarrow x : C_2$

Como $L(x, C_x) = P(x|C_x)$, temos

$$L(x, C_1) = \frac{1}{\sqrt{2\pi}} \exp\left[\frac{-(x+1)^2}{2}\right]$$

$$L(x, C_2) = \frac{1}{\sqrt{2\pi}} \exp\left[\frac{-(x-1)^2}{2}\right]$$

Encontrando o limiar x

$L(x, C_1) = L(x, C_2) \rightarrow (x+1)^2 = (x-1)^2$

$x^2 + 2x + 1 = x^2 - 2x + 1 \rightarrow \boxed{x = 0}$

Assim, temos as regiões de decisão

$$\boxed{R(x) = \begin{cases} C_1, & x < 0 \\ C_2, & x > 0 \end{cases}}$$

b) $P(C_1) = 0{,}7$ e $P(C_2) = 0{,}3$

- [illegible]

$$\frac{0{,}7}{\cancel{\sqrt{2\pi}}} \exp\left[\frac{-(x+1)^2}{2}\right] = \frac{0{,}3}{\cancel{\sqrt{2\pi}}} \exp\left[\frac{-(x-1)^2}{2}\right]$$

$$\frac{2\ln(7/3)}{\cancel{2}} - \frac{(x+1)^2}{\cancel{2}} = \frac{-(x-1)^2}{\cancel{2}}$$

$$2\ln(7/3) - \cancel{x^2} - 2x - \cancel{1} = -\cancel{x^2} + 2x - \cancel{1}$$

$$x = \frac{\ln(7/3)}{2} = 0{,}85$$

Dado o gráfico [illegible], temos:

$$R(x) = \begin{cases} C_1, & x < 0{,}85 \\ C_2, & x \geq 0{,}85 \end{cases}$$

EFC1 - Regressão Linear

Felippe Trigueiro Angelo - RA: 210479

26 de Setembro de 2019

Parte II - Atividade computacional

a) Regressão Linear utilizando a Equação Normal

No presente exercício será construído um preditor linear, onde com base em K amostras passadas será previsto o valor no instante de tempo imediatamente posterior. Assim, o vetor x(n) de entradas será constituído por x(n) = [x(n-1) ... x(n-K)]. Portanto, para cada valor de x(n) será calculado um valor y(n) correspondente a estimação.

Para este exercício será utilizado o conjunto de dados correspondente a temperatura mínima da cidade de Melbourne nos anos de 1981 até 1990.

O conjunto de dados de temperatura está disposto em 3650 linhas e 2 colunas, onde cada linha representa uma amostra de temperatura. As colunas correspondem às datas e as temperaturas mínimas. O conjunto de dados não contém ausência de nenhum valor.

Antes da realização do treinamento, o conjunto de dados foi separado em duas partes, o conjunto de treinamento e o conjunto de teste. O conjunto de treinamento é formado pelas temperaturas correspondentes aos anos de 1981 até 1989. Assim, o conjunto de teste é formado pelas temperaturas referentes ao ano de 1990.

Para o treinamento da primeira parte do exercício foi avaliado o algoritmo de regressão linear. A sua implementação foi realizada utilizando a equação normal, assim, a minimização da soma dos mínimos quadrados (SSD) foi utilizada como métrica. Não foi realizada a normalização nem a regularização dos dados.

$$\mathbf{X} = \begin{bmatrix} \mathbf{x_1} \\ \mathbf{x_2} \\ \vdots \\ \mathbf{x_N} \end{bmatrix} \qquad \mathbf{\Phi} = \mathbf{\Phi}(\mathbf{X}) = \begin{bmatrix} 1 & \mathbf{x_1} \\ 1 & \mathbf{x_2} \\ & \vdots \\ 1 & \mathbf{x_N} \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

Modelo:

$$\hat{\mathbf{y}} = \hat{\mathbf{y}}(\mathbf{X}) = \mathbf{\Phi}\mathbf{w} \tag{1}$$

$$\mathbf{w} = (\mathbf{\Phi}^T\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{y} \tag{2}$$

Para a seleção da quantidades de amostras passadas (K) foi utilizado o esquema de validação cruzada utilizando a abordagem k-fold. Nesse caso foi utilizado um valor de k igual a 4, correspondendo assim a 4 divisões do conjunto de treinamento.
Para avaliar qual o melhor valor de K foi calculada a raiz da soma dos quadrados (RMSE) no conjunto de validação, para os valores de K no intervalo de 1 a 30. Assim, o valor de K a ser escolhido corresponde a quantidade de amostras passadas que resulta no menor valor de erro. O gráfico correspondente ao desempenho do erro RMS com relação ao número de amostras passadas (K) pode ser visto na Figura 1. No processo de validação cruzada, pode ser visto que o menor valor de K (2.4291) corresponde a 18 amostras passadas.
Por fim, o modelo foi avaliado no conjunto de testes e a comparação pode ser vista no gráfico da Figura 2. O erro RMS avaliado no conjunto de testes foi de 4.95.

b) Regressão Linear utilizando Ridge Regression

Na segunda parte do exercício, o conjunto de amostras utilizadas para o treinamento foi obtido por meio da equação $\mathbf{X}'(n) = tanh(\mathbf{w}_k\mathbf{x}(n))$, onde os coeficientes w são obtidos aleatoriamente por meio de uma distribuição normal.
Na formação do conjunto de treinamento foram avaliados amostras com tamanhos (T) que variam de 1 até 100. Assim, para cada valor de T são obtidos, T vetores w. Além disso, cada vetor w tinha tamanho 5, assim o algoritmo de aprendizado recebe como entrada um conjunto de vetores contendo 5 amostras passadas.

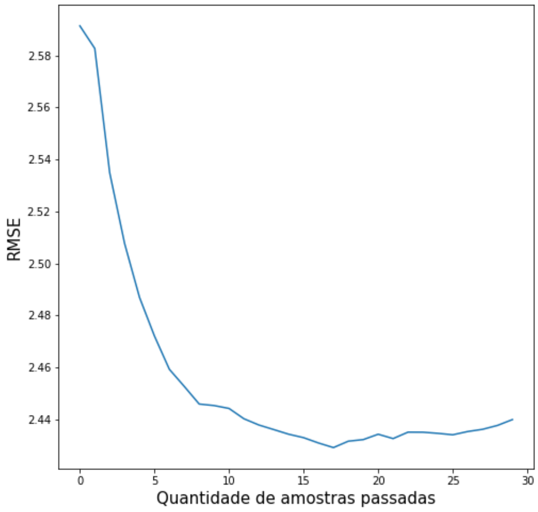


Figura 1: Desempenho do erro RMS com relação ao número de amostras passadas. O desempenho foi obtido por meio do algoritmo de regressão linear sem regularização. O menor valor de erro foi 2.4291 e foi obtido para K igual a 18.

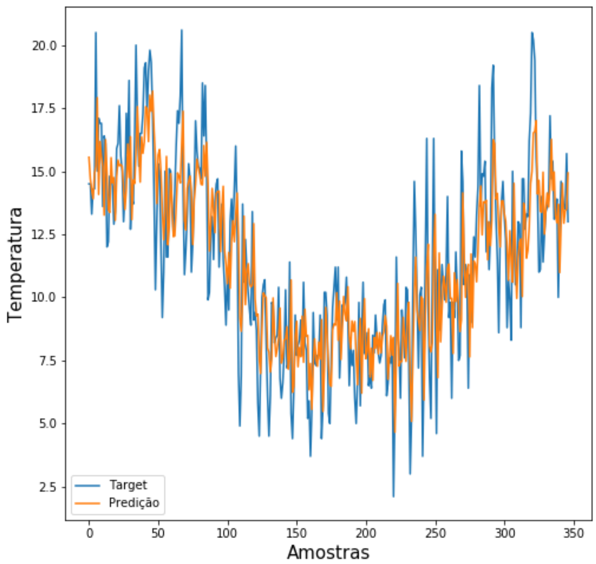


Figura 2: Comparação dos dados do conjunto de teste (Target) com os dados obtidos na predição.

A função tangente hiperbólico possui um comportamento aproximadamente linear nas regiões de -1 a 1, e um comportamento não linear nas outras regiões, juntamente com uma saturação para valores maiores que 3 (ou menores que -3). Assim, para evitar que o conjunto de dados seja composto de valores saturados, se faz necessário ajustar a escala dos valores x(n) bem como dos valores dos coeficientes w. Para o presente exercício, os valores de w foram obtidos na faixa entre 0 e 0.45 e os valores de x(n) foram normalizados para a faixa entre 0 e 1. Assim, os valores de x' estarão na faixa de 0 e 2.25.

Como modelo de treinamento foi utilizado a regressão linear juntamente com a regularização com ridge regression. Assim, o vetor de coeficientes pode ser obtido por meio da equação normal abaixo. Para a regularização foram inicialmente testados valores de lambda entre 0 e 1 com um passo de 0.01. Como foi observado que para um dado T, os valores de Erro RMS aumentavam, o treinamento foi realizado apenas para essa faixa.

Para a obtenção dos valores de T e de lambda, foi utilizado a abordagem da validação cruzada k-fold com um valor de k igual a 4. Assim, os valores de T e lambda que minimizaram o erro RMS foram utilizados na construção do modelo.

$$\mathbf{w} = (\mathbf{\Phi}^T \mathbf{\Phi} + \lambda \mathbf{I'})^{-1} \mathbf{\Phi}^T \mathbf{y} \tag{3}$$

Para a análise do erro RMS com relação ao tamanho do vetor de entrada (T), foi obtido o gráfico do desempenho do erro RMS com relação a T. Para a sua obtenção, para um dado valor de T foi calculado a média do erro RMS entre todos os valores de regularização. O resultado pode ser visto na Figura 3.

Além disso, também foi obtido o valor de lambda que minimizou o erro RMS para um dado valor de T. Os valores podem ser vistos na Figura 4.

Fazendo a análise do menor erro RMS obtido para um dado T e Lambada no conjunto de validação, foi obtido o valor de 2.4687, obtido para T igual a 11 e lambda igual a 0. Assim, com esses valores em mãos, foi obtido um novo modelo e o seu desempenho foi avaliado no conjunto de teste. A comparação entre os dados de teste e os dados obtidos no modelo pode ser vista na Figura 5. O valor do erro RMS obtido para esse conjunto de dados foi de 5.3753.

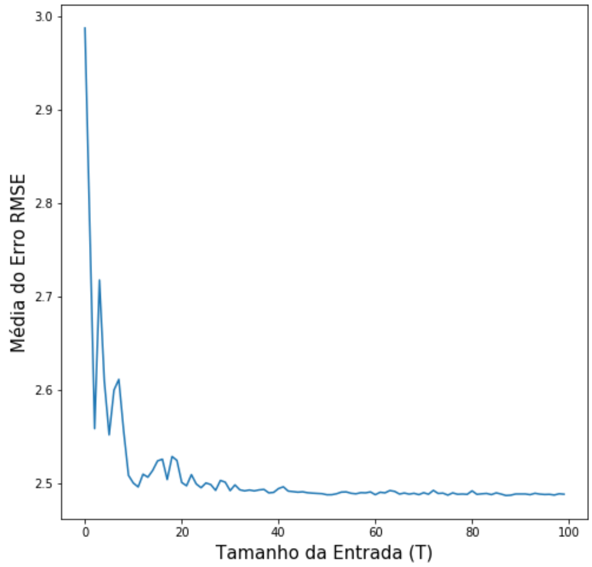


Figura 3: Média do erro RMS para os valores de regularização, em função do tamanho do vetor de entrada (T).

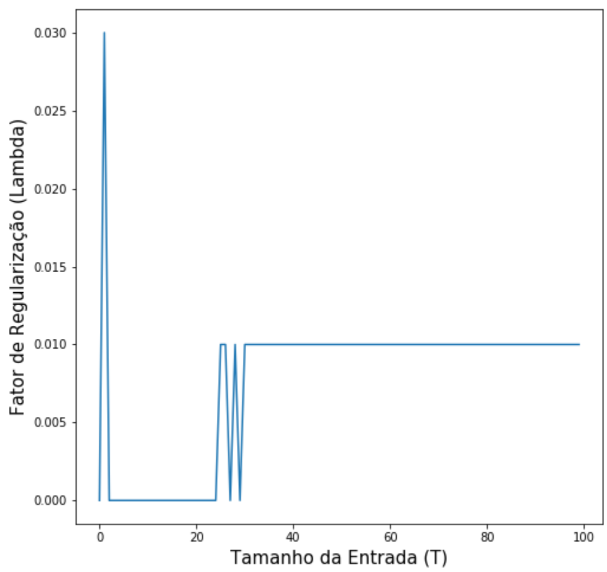


Figura 4: Fator de regularização (lambda) em função do tamanho do vetor de entrada (T).

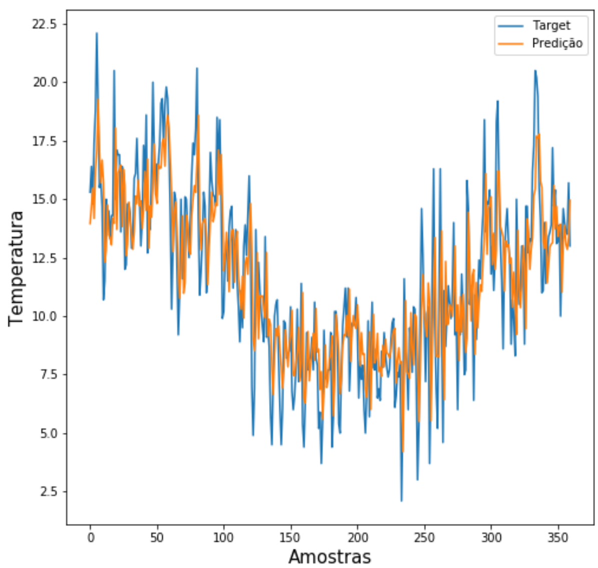


Figura 5: Comparação dos dados do conjunto de teste (Target) com os dados obtidos na predição por meio dos dados não lineares.