Exercícios Deep Learning - List Teórica 01
Elaboração de Carolina Coimbra, Guilherme Borges e Edemir Andrade Jr.

August 23, 2019

1 Retas

1- Esboce num gráfico as seguintes retas:

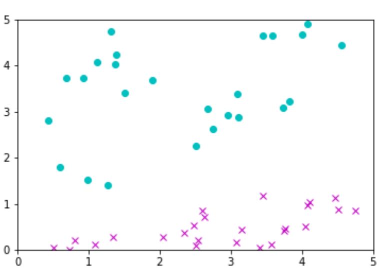
a) $2x_2 + x_1 = 0$

b) $x_2 - 2x_1 + 1 = 0$

c) $x_2 - 1 = 0$

d) $x_1 - 1 = 0$

2- Especifique uma reta que divide as duas categorias de itens no gráfico abaixo onde x_1 é o eixo abscissas e e x_2 é o eixo das ordenadas.



3- Considere a reta $x_2 = 3 + 2x_1$. Obtenha a expressão analítica do conjunto de todas as retas paralelas e o conjunto de todas as retas perpendiculares à reta acima.

2 Álgebra Linear

4- Sejam $w = w_1, ...w_n$ e $x = x_1, ..., x_n$ vetores coluna de dimensão $n \times 1$. Expresse $w'x$ em termos de um somatório.

5- Seja $x = (x_1, ..., x_n)$ um vetor-coluna $n \times 1$ e A uma matriz $n \times n$. A' indica a matriz transposta de A. Verifique que as seguintes identidades matriciais estão corretas, checando se o lado direito é igual ao lado esquerdo.

a) $x'Ax = \sum_{i,j} x_i x_j A_{ij}$

b) $x'x = \sum_i x_i^2$

c) xx' é uma matriz simétrica $n \times n$ com elemento (i, j) dado por $x_i x_j$

3 Derivadas

6- Encontre a derivada $F'(x)$ de

$$F(x) = \sqrt{x^2 + 1}$$

7- Encontre a derivada $F'(x)$ de

$$F(x) = e^{\sin x}$$

8- Função sigmóide:

$$S(x) = \frac{1}{1 + e^{-x}}$$

a) Esboce o gráfico de $S(x)$.

b) Mostre que $S(-x) = 1 - S(x)$

c) Calcule a derivada em termos da própria sigmóide, isto é, mostre que $S'(x) = S(x)(1 - S(x))$. Esboce o gráfico da derivada.

d) Qual o valor máximo de $S'(x)$? Para qual valor de x ela atinge esse máximo?

e) Considere $S(z) = \frac{1}{1+e^{-z}}$, sendo que $z = b + W_1 x$ encontre $\frac{\partial S}{\partial b}$ e $\frac{\partial S}{\partial W_1}$.

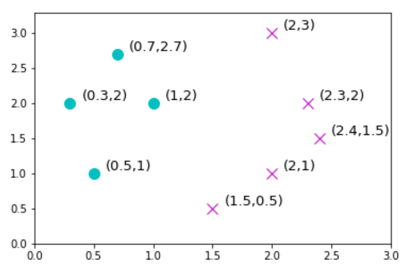
f) Considere $S(h(x)) = \frac{1}{1+e^{-h(x)}}$, calcule $\frac{\partial S}{\partial x}$ em função de $h(x)$.

9- Suponha que você tenha dados da forma $\mathbf{X} = (X_1, X_2, ..., X_n)$, onde $X_i \in \mathbb{R}$ e que seu classificador seja da forma $\hat{Y}_i = \beta X_i$. Considerando o erro quadrático, ou seja, $L(\hat{Y}, Y) = \sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2$, qual o valor de β que minimiza o erro?

4 Perceptron

10- Considere um perceptron com duas features x_1, x_2, onde $w_0 = 3, w_1 = 2, w_2 = 1$, qual é a fórmula da reta que divide as duas classes? Qual é o vetor normal à reta?

11- Considere um perceptron com $w_0 = 0, w_1 = -1, w_2 = 1$, com os pontos mostrados no gráfico abaixo (círculos pertencem à classe 1 e 'x' à classe 0), esboce o gráfico da reta de separação das classes. Execute uma iteração do algoritmo do perceptron com $m = 0.1$, esboce a nova reta de separação.



12- Seja o conjunto de entrada dado por um total de 4 amostras, onde cada amostra é representada pela tupla $(\boldsymbol{x_i}, \boldsymbol{t})$, composta pelo vetor $\boldsymbol{x_i} = (x_0, x_1, x_2)$ e um rotulo $\boldsymbol{t}$ associado a amostra.

	x_0	x_1	x_2	t
Entrada 1	1	0	0	0
Entrada 2	1	0	1	0
Entrada 3	1	1	0	0
Entrada 4	1	1	1	1

a) Execute a quinta iteração do algoritmo do perceptron com pesos iniciais $w_0, w_1 e w_2$ iguais a 0, taxa de aprendizado η igual a 0.5, e utilizando a função de ativação degrau bipolar definida como:

$$f(x) = \begin{cases} -1 & \text{se } x < 0 \\ +1 & \text{se } x \geq 0 \end{cases}$$

Abaixo segue o exemplo das quatro primeiras iterações do algoritmo:

1º Iteração:
Entrada 1:

$$\begin{aligned} s_{out} &= f(w_0x_0 + w_1x_1 + w_2x_2) \\ &= f(0*1+0*0+0*0) = f(0) = 0;\ logo\ s_{out} = t \end{aligned} \tag{1}$$

Entrada 2:

$$\begin{aligned} s_{out} &= f(w_0x_0 + w_1x_1 + w_2x_2) \\ &= f(0*1+0*0+0*1) = f(0) = 0;\ logo\ s_{out} = t \end{aligned} \tag{2}$$

Entrada 3:

$$\begin{aligned} s_{out} &= f(w_0x_0 + w_1x_1 + w_2x_2) \\ &= f(0*1+0*1+0*0) = f(0) = 0;\ logo\ s_{out} = t \end{aligned} \tag{3}$$

Entrada 4:

$$\begin{aligned} s_{out} &= f(w_0x_0 + w_1x_1 + w_2x_2) \\ &= f(0*1+0*1+0*1) = f(0) = 0;\ logo\ s_{out} \neq t \end{aligned} \tag{4}$$

Atualiza Pesos:

$$\begin{aligned} w_0 &= w_0 + \eta(t - s_{out}) * x_0 = 0 + 0.5(1-0)*1 = 0.5 \\ w_1 &= w_1 + \eta(t - s_{out}) * x_1 = 0 + 0.5(1-0)*1 = 0.5 \\ w_2 &= w_2 + \eta(t - s_{out}) * x_2 = 0 + 0.5(1-0)*1 = 0.5 \end{aligned}$$

2º Iteração:
Entrada 1:

$$\begin{aligned} s_{out} &= f(w_0x_0 + w_1x_1 + w_2x_2) \\ &= f(0.5*1+0.5*0+0.5*0) = f(0.5) = 1;\ logo\ s_{out} \neq t \end{aligned} \tag{5}$$

Atualiza Pesos:

$$\begin{aligned} w_0 &= w_0 + \eta(t - s_{out}) * x_0 = 0.5 + 0.5(0-1)*1 = 0 \\ w_1 &= w_1 + \eta(t - s_{out}) * x_1 = 0.5 + 0.5(0-1)*0 = 0.5 \\ w_2 &= w_2 + \eta(t - s_{out}) * x_2 = 0.5 + 0.5(0-1)*0 = 0.5 \end{aligned}$$

Entrada 2:

$$\begin{aligned} s_{out} &= f(w_0x_0 + w_1x_1 + w_2x_2) \\ &= f(0*1+0.5*0+0.5*1) = f(0.5) = 1;\ logo\ s_{out} \neq t \end{aligned} \tag{6}$$

Atualiza Pesos:

$$\begin{aligned} w_0 &= w_0 + \eta(t - s_{out}) * x_0 = 0 + 0.5(0-1)*1 = -0.5 \\ w_1 &= w_1 + \eta(t - s_{out}) * x_1 = 0.5 + 0.5(0-1)*0 = 0.5 \\ w_2 &= w_2 + \eta(t - s_{out}) * x_2 = 0.5 + 0.5(0-1)*1 = 0 \end{aligned}$$

Entrada 3:

$$\begin{aligned} s_{out} &= f(w_0x_0 + w_1x_1 + w_2x_2) \\ &= f(-0.5 * 1 + 0.5 * 1 + 0 * 0) = f(0) = 0;\ logo\ s_{out} = t \end{aligned} \tag{7}$$

Entrada 4:

$$\begin{aligned} s_{out} &= f(w_0x_0 + w_1x_1 + w_2x_2) \\ &= f(-0.5 * 1 + 0.5 * 1 + 0 * 1) = f(0) = 0;\ logo\ s_{out} \neq t \end{aligned} \tag{8}$$

Atualiza Pesos:

$$\begin{aligned} w_0 &= w_0 + \eta(t - s_{out}) * x_0 = -0.5 + 0.5(1 - 0) * 1 = 0 \\ w_1 &= w_1 + \eta(t - s_{out}) * x_1 = 0.5 + 0.5(1 - 0) * 1 = 1 \\ w_2 &= w_2 + \eta(t - s_{out}) * x_2 = 0 + 0.5(1 - 0) * 1 = 0.5 \end{aligned}$$

3º Iteração:
Entrada 1:

$$\begin{aligned} s_{out} &= f(w_0x_0 + w_1x_1 + w_2x_2) \\ &= f(0 * 1 + 1 * 0 + 0.5 * 0) = f(0) = 0;\ logo\ s_{out} = t \end{aligned} \tag{9}$$

Entrada 2:

$$\begin{aligned} s_{out} &= f(w_0x_0 + w_1x_1 + w_2x_2) \\ &= f(0 * 1 + 1 * 0 + 0.5 * 1) = f(0.5) = 1;\ logo\ s_{out} \neq t \end{aligned} \tag{10}$$

Atualiza Pesos:

$$\begin{aligned} w_0 &= w_0 + \eta(t - s_{out}) * x_0 = 0 + 0.5(0 - 1) * 1 = -0.5 \\ w_1 &= w_1 + \eta(t - s_{out}) * x_1 = 1 + 0.5(0 - 1) * 0 = 1 \\ w_2 &= w_2 + \eta(t - s_{out}) * x_2 = 0.5 + 0.5(0 - 1) * 1 = 0 \end{aligned}$$

Entrada 3:

$$\begin{aligned} s_{out} &= f(w_0x_0 + w_1x_1 + w_2x_2) \\ &= f(-0.5 * 1 + 1 * 1 + 0 * 0) = f(0.5) = 1;\ logo\ s_{out} \neq= t \end{aligned} \tag{11}$$

Atualiza Pesos:

$$\begin{aligned} w_0 &= w_0 + \eta(t - s_{out}) * x_0 = -0.5 + 0.5(0 - 1) * 1 = -1 \\ w_1 &= w_1 + \eta(t - s_{out}) * x_1 = 1 + 0.5(0 - 1) * 0 = 1 \\ w_2 &= w_2 + \eta(t - s_{out}) * x_2 = 0.5 + 0.5(0 - 1) * 1 = 0 \end{aligned}$$

Entrada 4:

$$\begin{aligned} s_{out} &= f(w_0x_0 + w_1x_1 + w_2x_2) \\ &= f(-1 * 1 + 1 * 1 + 0 * 1) = f(0) = 0;\ logo\ s_{out} \neq t \end{aligned} \tag{12}$$

Atualiza Pesos:

$$
\begin{aligned}
w_0 &= w_0 + \eta(t - s_{out}) * x_0 = -1 + 0.5(1-0) * 1 = -0.5 \\
w_1 &= w_1 + \eta(t - s_{out}) * x_1 = 1 + 0.5(1-0) * 1 = 1.5 \\
w_2 &= w_2 + \eta(t - s_{out}) * x_2 = 0 + 0.5(1-0) * 1 = 0.5
\end{aligned}
$$

4^o Iteração:
Entrada 1:

$$
\begin{aligned}
s_{out} &= f(w_0x_0 + w_1x_1 + w_2x_2) \\
&= f(-0.5 * 1 + 1.5 * 0 + 0.5 * 0) = f(-0.5) = 0;\ logo\ s_{out} = t
\end{aligned} \tag{13}
$$

Entrada 2:

$$
\begin{aligned}
s_{out} &= f(w_0x_0 + w_1x_1 + w_2x_2) \\
&= f(-0.5 * 1 + 1.5 * 0 + 0.5 * 1) = f(0) = 0;\ logo\ s_{out} = t
\end{aligned} \tag{14}
$$

Entrada 3:

$$
\begin{aligned}
s_{out} &= f(w_0x_0 + w_1x_1 + w_2x_2) \\
&= f(-0.5 * 1 + 1.5 * 1 + 0.5 * 0) = f(1) = 1;\ logo\ s_{out} \neq= t
\end{aligned} \tag{15}
$$

Atualiza Pesos:

$$
\begin{aligned}
w_0 &= w_0 + \eta(t - s_{out}) * x_0 = -0.5 + 0.5(0-1) * 1 = -1 \\
w_1 &= w_1 + \eta(t - s_{out}) * x_1 = 1.5 + 0.5(0-1) * 0 = 1 \\
w_2 &= w_2 + \eta(t - s_{out}) * x_2 = 0.5 + 0.5(0-1) * 1 = 0.5
\end{aligned}
$$

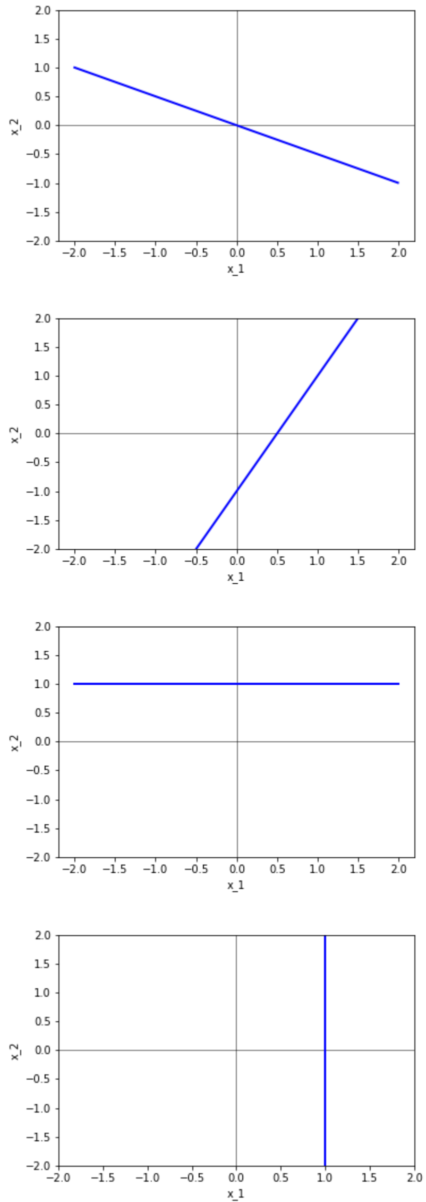
Entrada 4:

$$
\begin{aligned}
s_{out} &= f(w_0x_0 + w_1x_1 + w_2x_2) \\
&= f(-1 * 1 + 1 * 1 + 0.5 * 1) = f(0.5) = 1;\ logo\ s_{out} = t
\end{aligned} \tag{16}
$$

b) Esboce o gráfico da reta gerada após a quinta iteração do algoritmo. A equação da reta após a quinta iteração é dado pela equação: $x_1w_1 + x_2w_2 = -w_0$

Solução

1-



2- $x_2 - \frac{1}{2}x_1 = 0$

3- Paralelas : $x_2 = 2x_1 + c$ com $c \in \mathbb{R}$
Perpendiculares: $x_2 = -\frac{1}{2}x_1 + c$ com $c \in \mathbb{R}$

4- $\sum_{i=1}^{n} w_i x_i$

5- a)

$$\begin{aligned}\mathbf{x'Ax} &= \begin{bmatrix}\sum_j x_1 A_{1,j} & \cdots & \sum_j x_n A_{n,j}\end{bmatrix}\mathbf{x} \\ &= \sum_{i,j} x_i A_{i,j} x_j\end{aligned}$$

b)

$$\begin{aligned}\mathbf{x'x} &= \begin{bmatrix}x_1 & \cdots & x_n\end{bmatrix}\begin{bmatrix}x_1 \\ \vdots \\ x_n\end{bmatrix} \\ &= \sum_i x_i^2\end{aligned}$$

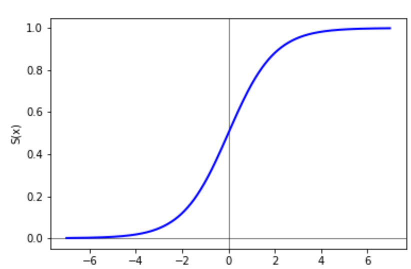
c)

$$\begin{aligned}\mathbf{xx'} &= \begin{bmatrix}x_1 \\ \vdots \\ x_n\end{bmatrix}\begin{bmatrix}x_1 & \cdots & x_n\end{bmatrix} \\ &= \begin{bmatrix}x_1x_1 & x_1x_2 & \cdots & x_1x_n \\ x_2x_1 & x_2x_2 & \cdots & x_2x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_nx_1 & x_nx_2 & \cdots & x_nx_n\end{bmatrix}\end{aligned}$$

6- $x(x^2+1)^{-\frac{1}{2}}$

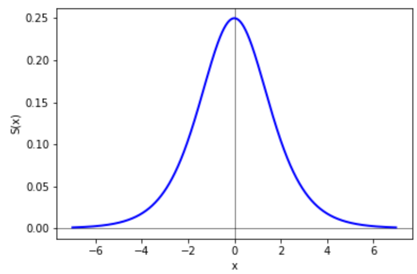
7- $\cos x e^{\sin x}$

8- a)



b) $S(-x) = \frac{1}{1+e^{x}} = \frac{1}{1+\frac{1}{e^{-x}}} = \frac{1}{\frac{e^{-x}+1}{e^{-x}}} = \frac{e^{-x}}{e^{-x}+1} = 1 - \frac{1}{1+e^{-x}} = 1 - S(x)$

c) $S'(x) = \frac{e^{-x}}{(1+e^{-x})^2} = \frac{1}{1+e^{-x}}\frac{e^{-x}}{1+e^{-x}} = S(x)(1-S(x))$



d) Valor máximo é 0.25 quando $x = 0$.

e) $\frac{\partial S}{\partial b} = \frac{e^{-(b+W_1x)}}{(1+e^{-(b+W_1x)})^2}$ e $\frac{\partial S}{\partial W_1} = \frac{xe^{-(b+W_1x)}}{(1+e^{-(b+W_1x)})^2}$

f) $\frac{\partial S}{\partial x} = \frac{h'(x)e^{-h(x)}}{(1+e^{-h(x)})^2}$

9- Para encontrar o valor mínimo da perda, devemos derivar a função em relação a β e encontrar onde ela é 0. $L(\hat{Y}, Y) = \sum_{i=1}^{n}(\beta X_i - Y_i)^2$

$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{n} 2X_i(\beta X_i - Y_i) = 2(\sum_{i=1}^{n}\beta X_i^2 - \sum_{i=1}^{n} X_iY_i) = 0$

$\beta = \frac{\sum_{i=1}^{n} X_iY_i}{\sum_{i=1}^{n} X_i^2}$

10- Reta: $x_2 = -3 - 2x_1$ Vetor normal: (2 1)

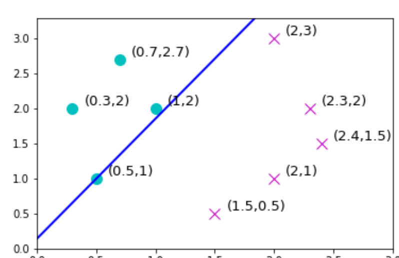
11- Reta: $x_2 = x_1$

Algoritmo:

O único ponto onde $y \neq \hat{y}$ é $(2, 3)$

$$\begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} + 0.1(-1)\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -0.1 \\ -1.2 \\ 0.7 \end{bmatrix}$$

Nova reta : $x_2 = \frac{1}{7} + \frac{12}{7}x_1$



12-

a)

5º Iteração:

Entrada 1:

$$\begin{aligned} s_{out} &= f(w_0x_0 + w_1x_1 + w_2x_2) \\ &= f(-1*1 + 1*0 + 0.5*0) = f(-1) = 0;\ logo\ s_{out} = t \end{aligned} \tag{17}$$

Entrada 2:

$$\begin{aligned} s_{out} &= f(w_0x_0 + w_1x_1 + w_2x_2) \\ &= f(-1*1 + 1*0 + 0.5*1) = f(-0.5) = 0;\ logo\ s_{out} = t \end{aligned} \tag{18}$$

Entrada 3:

$$\begin{aligned} s_{out} &= f(w_0x_0 + w_1x_1 + w_2x_2) \\ &= f(-1*1 + 1*1 + 0.5*0) = f(0) = 0;\ logo\ s_{out} = t \end{aligned} \tag{19}$$

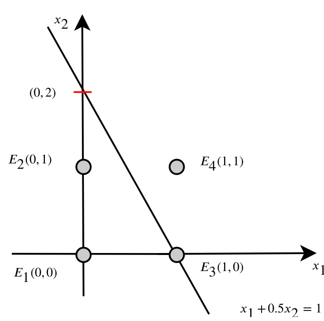
Entrada 4:

$$\begin{aligned} s_{out} &= f(w_0x_0 + w_1x_1 + w_2x_2) \\ &= f(-1*1 + 1*1 + 0.5*1) = f(0.5) = 1;\ logo\ s_{out} = t \end{aligned} \tag{20}$$

Logo temos o resultado: $w_0 = -1, w_1 = 1, w_2 = 0.5$

b)

Esboçando a reta $x_1 * 1 + x_2 * 0.5 = 1$ temos:



5 Derivadas

13- Seja $x, y \in \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^n$ e $\mathbf{y} \in \mathbb{R}^m$ de forma que:

$$\frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix} \qquad \frac{\partial \mathbf{y}}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} & \frac{\partial y_2}{\partial x} & \cdots & \frac{\partial y_m}{\partial x} \end{bmatrix}$$

e

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \mathbf{y}}{\partial x_1} \\ \frac{\partial \mathbf{y}}{\partial x_2} \\ \vdots \\ \frac{\partial \mathbf{y}}{\partial x_3} \end{bmatrix} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} \frac{\partial y_2}{\partial x_1} \cdots \frac{\partial y_m}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} \frac{\partial y_2}{\partial x_2} \cdots \frac{\partial y_m}{\partial x_2} \\ \vdots \\ \frac{\partial y_1}{\partial x_n} \frac{\partial y_2}{\partial x_n} \cdots \frac{\partial y_m}{\partial x_n} \end{bmatrix}$$

Assuma que $\mathbf{y} = f(\mathbf{u})$ e $\mathbf{u} = g(\mathbf{x})$, escreva a derivada (usando a regra da cadeia) para $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$

14- Seja $\mathbf{X} \in \mathbb{R}^{m \times n},\ \mathbf{w} \in \mathbb{R}^n,\ \mathbf{y} \in \mathbb{R}^m$ e $z = \|\mathbf{X}\mathbf{w} - \mathbf{y}\|^2$, calcule $\frac{\partial z}{\partial \mathbf{w}}$.

15- Sejam $u, \mathbf{x} \in R^n$ (vetores colunas).

a) Calcule a derivada de $u'\mathbf{x}$ em respeito a $\mathbf{x}$, ou seja: $\frac{\partial (u'\mathbf{x})}{\partial \mathbf{x}}$

b) Calcule a derivada de $\mathbf{x}'\mathbf{x}$ em respeito a $\mathbf{x}$, ou seja: $\frac{\partial (\mathbf{x}'\mathbf{x})}{\partial \mathbf{x}}$

6 Regressão Linear

16- Em uma regressão linear, o valor estimado $\hat{Y}_i$ é dado por

$$\hat{Y}_i = b + w_1 x_1 + ... + w_n x_n$$

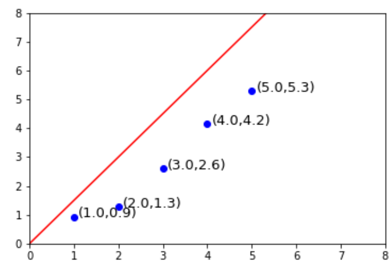
onde n é o número de features. O erro quadrático é dado por $L(\hat{Y}, Y) = \sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2$. Neste exercício, $n = 1$, portanto $\hat{Y}_i = b + w_1 x_1$.

a) Considerando os pontos mostrados no gráfico abaixo e a reta com $w_1 = 1.5$ e $b = 0$, calcule o erro quadrático dessa reta.

b) Utilize a derivada do erro quadrático para atualizar os valores de w_1 e b, encontrando uma nova reta (Dica: use uma taxa de aprendizado menor que 0.1). Qual é o erro quadrático desta nova reta?

7 Regressão Logística

17- No modelo de regressão logística com um regressor apenas: $P(Y_i = 1) = p(x_i) = \frac{1}{1+\exp(-b-w_1 x_i)}$. Deduza que a log-verossimilhança de $\theta = (b, w_1)$ no



caso do modelo logístico com um único regressor é dada por:

$$l(\theta) = b\sum_{i=1}^{n} y_i + w_1 \sum_{i=1}^{n} x_i y_i - \sum_{i} \log(1 + e^{b + w_1 x_i})$$

18- No modelo logístico com um regressor apenas, mostre que o vetor gradiente de $l(\theta)$ é dado por:

$$Dl(\theta) = \begin{bmatrix} \frac{\partial logl}{\partial b} \\ \frac{\partial logl}{\partial w_1} \end{bmatrix} \quad = \begin{bmatrix} \sum_{i=1}^{n} y_i - p_i \\ \sum_{i=1}^{n} x_i y_i - p_i x_i \end{bmatrix}$$

onde $p_i = p(x_i)$.

19- Ainda no modelo logístico com um regressor apenas, mostre que a matriz hessiana de $l(\theta)$ é dada por:

$$Dl(\theta) = -\begin{bmatrix} \frac{\partial^2 logl}{\partial b^2} & \frac{\partial^2 logl}{\partial b \partial w_1} \\ \frac{\partial^2 logl}{\partial b \partial w_1} & \frac{\partial^2 logl}{\partial w_1^2} \end{bmatrix} \quad = -\begin{bmatrix} \sum_{i=1}^{n} p_i(1-p_i) & \sum_{i=1}^{n} p_i(1-p_i)x_i \\ \sum_{i=1}^{n} p_i(1-p_i)x_i & \sum_{i=1}^{n} p_i(1-p_i)x_i^2 \end{bmatrix}$$

8 Newton e SGA

20- Aplique duas iterações do método de newton para encontrar o ponto máximo da função $f(x) = -(x-3)^4$ para $x_0 = 1$.

21- Utilizando a a mesma função do exercicio anterior, aplique o método do gradiente ascendente usando learning rate $\alpha = 0.01$ e $x_0 = 1$. Compare os dois resultados.

Solução

1- $\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \frac{\partial \mathbf{y}}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial \mathbf{x}}$

2-

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \quad \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

$$\mathbf{Xw \text{ - } Y} = \begin{bmatrix} \sum_i x_{i1} w_i - y_1 \\ \sum_i x_{i2} w_i - y_2 \\ \vdots \\ \sum_i x_{im} w_i - y_m \end{bmatrix}$$

$z = \|\mathbf{Xw} - \mathbf{y}\|^2$

$$\frac{\partial z}{\partial \mathbf{w}} = \frac{\partial \sum_{i,j} (x_{ij} w_i - y_j)^2}{\partial \mathbf{w}} = \begin{bmatrix} \frac{\partial z}{\partial w_1} \\ \frac{\partial z}{\partial w_2} \\ \vdots \\ \frac{\partial z}{\partial w_n} \end{bmatrix} \quad \begin{bmatrix} 2 \sum_{ij} (x_{ij} w_i - y_j) \sum_j x_{1j} \\ 2 \sum_{ij} (x_{ij} w_i - y_j) \sum_j x_{2j} \\ \vdots \\ 2 \sum_{ij} (x_{ij} w_i - y_j) \sum_j x_{mj} \end{bmatrix}$$

3-

a) A derivada de $u'\mathbf{x} = \sum_{i=1}^{n} u_i x_i$ em respeito a $\mathbf{x}$:

$$\frac{\partial \sum_{i=1}^{n} u_i x_i}{\partial x_i} = u_i \Rightarrow \frac{\partial u'\mathbf{x}}{\partial \mathbf{x}} = (u_1, \ldots, u_n) = u'$$

b) A derivada de $\mathbf{x}'\mathbf{x} = \sum_{i=1}^{n} x_i^2$ em respeito a $\mathbf{x}$:

$$\frac{\partial \sum_{i=1}^{n} x_i^2}{\partial x_i} = 2x_i \Rightarrow \frac{\partial \mathbf{x}'\mathbf{x}}{\partial \mathbf{x}} = (2\mathbf{x}_1, \ldots, 2\mathbf{x}_n) = 2\mathbf{x}'$$

4-

a)

$(1.5 - 0.9)^2 + (3 - 1.3)^2 + (4.5 - 2.6)^2 + (6 - 4.2)^2 + (7.5 - 5.3)^2 = 14.94$

b)

$$\frac{\partial L}{\partial b} = 2\sum_{i=1}^{n}(b + w_1 x_i - Y_i) = 2nb + 2\sum_{i=1}^{n}(w_1 x_i - Y_i)$$

$$\frac{\partial L}{\partial w_1} = 2\sum_{i=1}^{n}\big((b + w_1 x_i - Y_i)x_i\big) = 2b\sum_{i=1}^{n} x_i + 2\sum_{i=1}^{n}(w_1 x_i^2 - Y_i x_i)$$

Colocando a taxa de aprendizado $\alpha = 0.01$

$$\begin{bmatrix} w^* \\ b^* \end{bmatrix} = \begin{bmatrix} 1.5 \\ 0 \end{bmatrix} - \alpha \begin{bmatrix} 2\sum_{i=1}^{n}(1.5x_i^2 - Y_i x_i) \\ 2\sum_{i=1}^{n}(1.5x_i - Y_i) \end{bmatrix} = \begin{bmatrix} 0.942 \\ -0.164 \end{bmatrix}$$

$$L = (0.78-0.9)^2 + (1.72-1.3)^2 + (2.66-2.6)^2 + (3.6-4.2)^2 + (4.55-5.3)^2 = 1.117$$

5-

$$
\begin{aligned}
l(\theta) &= log\left(\prod_{i=1}^{n} \mathbb{P}(Y_i = y_i)\right) \\
&= log\left(\prod_{i=1}^{n} \mathbb{P}(Y_i = y_i)\right) \\
&= log\left(\prod_{i=1}^{n} p(x_i)^{y_i}(1 - p(x_i))^{1-y_i}\right) \\
&= \sum_{i=1}^{n} log(p(x_i)^{y_i}(1 - p(x_i))^{1-y_i}) \\
&= \sum_{i=1}^{n} log(p(x_i)^{y_i}) + \sum_{i=1}^{n} log((1 - p(x_i))^{1-y_i}) \\
&= \sum_{i=1}^{n} y_i log(p(x_i)) + \sum_{i=1}^{n} (1 - y_i) log(1 - p(x_i)) \\
&= \sum_{i=1}^{n} y_i log\left(\frac{1}{1 + exp(-b - w_1 x_i)}\right) + \sum_{i=1}^{n} (1 - y_i) log\left(1 - \frac{1}{1 + exp(-b - w_1 x_i)}\right) \\
&= \sum_{i=1}^{n} y_i log\left(\frac{exp(b + w_1 x_i)}{1 + exp(b + w_1 x_i)}\right) + \sum_{i=1}^{n} (1 - y_i) log\left(\frac{1}{1 + exp(b + w_1 x_i)}\right) \\
&= \sum_{i=1}^{n} y_i log(exp(b + w_1 x_i)) - \sum_{i=1}^{n} y_i log(1 + exp(b + w_1 x_i)) + \sum_{i=1}^{n} (1 - y_i) log\left(\frac{1}{1 + exp(b + w_1 x_i)}\right) \\
&= \sum_{i=1}^{n} y_i log(exp(b + w_1 x_i)) - \sum_{i=1}^{n} y_i log(1 + exp(b + w_1 x_i)) - \sum_{i=1}^{n} (1 - y_i) log(1 + exp(b + w_1 x_i)) \\
&= \sum_{i=1}^{n} y_i (b + w_1 x_i) - \sum_{i=1}^{n} y_i log(1 + exp(b + w_1 x_i)) - \sum_{i=1}^{n} (1 - y_i) log(1 + exp(b + w_1 x_i)) \\
&= b \sum_{i=1}^{n} y_i + w_1 \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} y_i log(1 + exp(b + w_1 x_i)) - \sum_{i=1}^{n} (1 - y_i) log(1 + exp(b + w_1 x_i)) \\
&= b \sum_{i=1}^{n} y_i + w_1 \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} log(1 + exp(b + w_1 x_i))
\end{aligned}
$$

6-

$$\begin{aligned}
\frac{\partial l(\theta)}{\partial b} &= \frac{\partial}{\partial b}\left(b\sum_{i=1}^{n} y_i + w_1\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} log(1+exp(b+w_1x_i))\right)\\
&= \frac{\partial}{\partial b}\left(b\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} log(1+exp(b+w_1x_i))\right)\\
&= \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \frac{\partial}{\partial b}(log(1+exp(b+w_1x_i)))\\
&= \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \frac{\frac{\partial}{\partial b}(1+exp(b+w_1x_i))}{1+exp(b+w_1x_i)}\\
&= \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \frac{exp(b+w_1x_i)}{1+exp(b+w_1x_i)}\\
&= \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} p_i\\
&= \sum_{i=1}^{n}(y_i - p_i)\\
\frac{\partial l(\theta)}{\partial w_1} &= \frac{\partial}{\partial w_1}\left(b\sum_{i=1}^{n} y_i + w_1\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} log(1+exp(b+w_1x_i))\right)\\
&= \frac{\partial}{\partial w_1}\left(w_1\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} log(1+exp(b+w_1x_i))\right)\\
&= \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} \frac{\partial}{\partial w_1}(log(1+exp(b+w_1x_i)))\\
&= \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} \frac{\frac{\partial}{\partial w_1}(1+exp(b+w_1x_i))}{1+exp(b+w_1x_i)}\\
&= \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} \frac{x_i exp(b+w_1x_i)}{1+exp(b+w_1x_i)}\\
&= \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} x_i p_i\\
&= \sum_{i=1}^{n}(y_i x_i - x_i p_i)
\end{aligned}$$

7-

$$
\begin{aligned}
\frac{\partial^2 l(\theta)}{\partial b^2} &= \frac{\partial}{\partial b}\left(\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} p_i\right) \\
&= \frac{\partial}{\partial b}\left(\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \frac{1}{1+exp(-b-w_1x_i)}\right) \\
&= -\sum_{i=1}^{n} \frac{\partial}{\partial b}\left(\frac{1}{1+exp(-b-w_1x_i)}\right) \\
&= -\sum_{i=1}^{n} \left(-\frac{-exp(-b-w_1x_i)}{(1+exp(-b-w_1x_i))^2}\right) \\
&= -\sum_{i=1}^{n} \left(\frac{exp(-b-w_1x_i)}{(1+exp(-b-w_1x_i))^2}\right) \\
&= -\sum_{i=1}^{n} \left(\frac{exp(-b-w_1x_i)}{1+exp(-b-w_1x_i)}\frac{1}{1+exp(-b-w_1x_i)}\right) \\
&= -\sum_{i=1}^{n} \left(\frac{1}{1+exp(b+w_1x_i)}\frac{1}{1+exp(-b-w_1x_i)}\right) \\
&= -\sum_{i=1}^{n}(1-p_i)p_i
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 l(\theta)}{\partial b \partial w_1} = \frac{\partial^2 l(\theta)}{\partial w_1 \partial b} &= \frac{\partial}{\partial w_1}\left(\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} p_i\right) \\
&= \frac{\partial}{\partial w_1}\left(\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \frac{1}{1+exp(-b-w_1x_i)}\right) \\
&= -\sum_{i=1}^{n} \frac{\partial}{\partial w_1}\left(\frac{1}{1+exp(-b-w_1x_i)}\right) \\
&= -\sum_{i=1}^{n} \left(-\frac{-x_i exp(-b-w_1x_i)}{(1+exp(-b-w_1x_i))^2}\right) \\
&= -\sum_{i=1}^{n} \left(\frac{x_i exp(-b-w_1x_i)}{(1+exp(-b-w_1x_i))^2}\right) \\
&= -\sum_{i=1}^{n} \left(\frac{x_i exp(-b-w_1x_i)}{1+exp(-b-w_1x_i)} \frac{1}{1+exp(-b-w_1x_i)}\right) \\
&= -\sum_{i=1}^{n} \left(x_i \frac{1}{1+exp(b+w_1x_i)} \frac{1}{1+exp(-b-w_1x_i)}\right) \\
&= -\sum_{i=1}^{n} x_i(1-p_i)p_i
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 l(\theta)}{\partial w_1^2} &= \frac{\partial}{\partial w_1}\left(\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} x_i p_i\right) \\
&= \frac{\partial}{\partial w_1}\left(\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} x_i p_i\right) \\
&= \frac{\partial}{\partial w_1}\left(-\sum_{i=1}^{n} x_i p_i\right) \\
&= \frac{\partial}{\partial w_1}\left(-\sum_{i=1}^{n} x_i \frac{1}{1+exp(-b-w_1x_i)}\right) \\
&= -\sum_{i=1}^{n} x_i \frac{x_i exp(-b-w_1x_i)}{(1+exp(-b-w_1x_i))^2} \\
&= -\sum_{i=1}^{n} x_i x_i \frac{exp(-b-w_1x_i)}{1+exp(-b-w_1x_i)} \frac{1}{1+exp(-b-w_1x_i)} \\
&= -\sum_{i=1}^{n} x_i^2(1-p_i)p_i
\end{aligned}
$$

8- Jupyter notebook

9-
Seja $x_0 = 1$ e f(x) e suas derivadas:

$$\begin{aligned} f(x) &= -(x-3)^4 \\ f^{'}(x) &= -4(x-3)^3 \\ f^{''}(x) &= -12(x-3)^2 \end{aligned} \tag{21}$$

Utilizando método de Newton: $x_{n+1} = x_n - \frac{f^{'}(x_n)}{f^{''}(x_n)}$

- Para $n = 0$ e $x_0 = 1$
  $x_1 = x_0 - \frac{f^{'}(x_0)}{f^{''}(x_0)} = 1.66$
- Para $n = 1$ e $x_1 = 1.66$
  $x_2 = x_1 - \frac{f^{'}(x_1)}{f^{''}(x_1)} = 2.11$
- Para $n = 2$ e $x_1 = 2.11$
  $x_3 = x_2 - \frac{f^{'}(x_2)}{f^{''}(x_2)} = 2.407$

10-

Utilizando método do Gradient Ascendente: $x_{n+1} = x_n + \alpha * f^{'}(x_n)$, onde $\alpha = 0.01$

- Para $n = 0$ e $x_0 = 1$
  $x_1 = x_0 + \alpha * f^{'}(x_0) = 1.32$
- Para $n = 1$ e $x_1 = 1.32$
  $x_2 = x_1 + \alpha * f^{'}(x_1) = 1.509$
- Para $n = 2$ e $x_2 = 1.509$
  $x_3 = x_2 + \alpha * f^{'}(x_2) = 1.64$