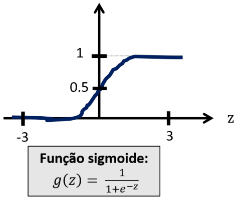
Fronteira de Decisão e Função Custo para Regressão Logística

Na aula anterior, revisamos o que são problemas de classificação e fizemos uma introdução sobre o método de Regressão Logística

Nesta aula continuaremos nossos estudos falando sobre **Fronteira de Decisão**, um conceito importante de ser bem compreendido no âmbito dos algoritmos de classificação.

Também vamos definir a **Função Custo** que utilizaremos para fins de Regressão Logística



Passo 1:

Definimos z como um modelo linear do tipo

$$z = \vec{w} \cdot \vec{x} + b$$

Passo 2:

Passamos este z pela função sigmoide

$$g(z) = \frac{1}{1 + e^{-z}}$$

Resultado

O resultado desse passo-a-passo é o modelo de Regressão Logística

$$f_{\vec{w},b}(\vec{x}) = g(\vec{w} \cdot \vec{x} + b) = \frac{1}{1 + e^{-(\vec{w} \cdot \vec{x} + b)}} \qquad \text{onde } 0 < f_{\vec{w},b}(\vec{x}) < 1$$

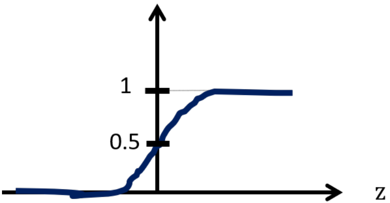
Gerando as previsões $\hat{y} = 0$ e $\hat{y} = 1$

Como $0 < f_{\overrightarrow{w},b}(\overrightarrow{x}) < 1$, podemos considerar que $f_{\overrightarrow{w},b}(\overrightarrow{x})$ é a probabilidade de que y seja 1.

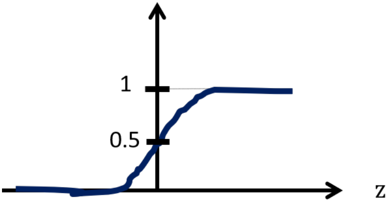
Ou seja, $f_{\overrightarrow{w},b}(\overrightarrow{x}) = P(y = 1|\overrightarrow{x}; \overrightarrow{w}, b)$

Sendo assim, podemos considerar que

$$\hat{y} = \begin{cases} 1 & \text{se } f_{\overrightarrow{w},b}(\overrightarrow{x}) \geq 0.5 \\ 0 & \text{se } f_{\overrightarrow{w},b}(\overrightarrow{x}) < 0.5 \end{cases}$$



Gerando as previsões $\hat{y} = 0$ e $\hat{y} = 1$



Pergunta:

Quando $f_{\vec{w},b}(\vec{x}) \geq 0.5$?

Resposta:

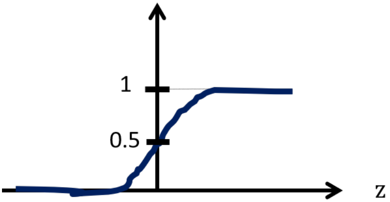
Quando $g(z) \geq 0.5$.

Ou, equivalentemente, quando $z \geq 0$

Ou, equivalentemente, quando $\vec{w} \cdot \vec{x} + b \geq 0$

Conclusão:

Nosso modelo fará a previsão $\hat{y} = 1$ quando $\vec{w} \cdot \vec{x} + b \geq 0$



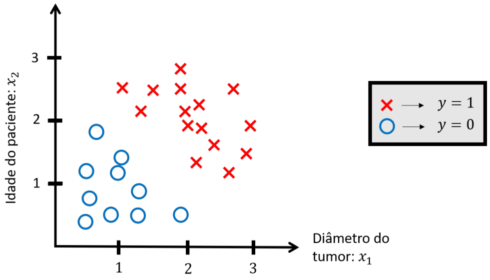
Analogamente, nosso modelo fará a previsão $\hat{y} = 0$ quando $\vec{w} \cdot \vec{x} + b < 0$

Definição:

Também podemos definir a fronteira de decisão, que consiste no valor de $\vec{x}$ que faz com que $z = \vec{w} \cdot \vec{x} + b = 0$.

Note que esse é o valor que faz com que $f_{\vec{w},b}(\vec{x}) = 0.5$

Seja o problema de classificação abaixo, que contém 2 variáveis.



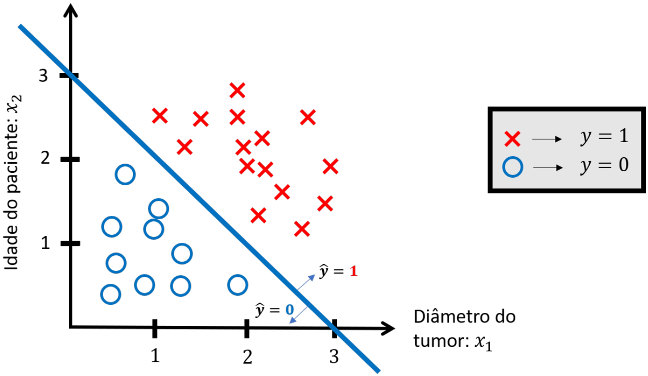
Como são 2 variáveis, o modelo para esse caso é

$$f_{\overrightarrow{w},b}(\overrightarrow{x}) = g(w_1 x_1 + w_2 x_2 + b)$$

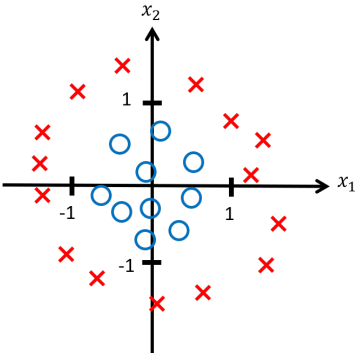
Supondo $w_1 = w_2 = 1$ e $b = -3$, temos a seguinte fronteira de decisão:

$$z = \overrightarrow{w} \cdot \overrightarrow{x} + b = 0 \quad \rightarrow \quad x_1 + x_2 - 3 = 0 \quad \rightarrow \quad x_1 + x_2 = 3$$

Ilustrando a fronteira de decisão $x_1 + x_2 = 3$ na figura, temos



Podemos ter também Fronteiras de Decisão mais complexas (não lineares):



Usando **Engenharia de Características**, podemos estabelecer o seguinte modelo para esse caso:

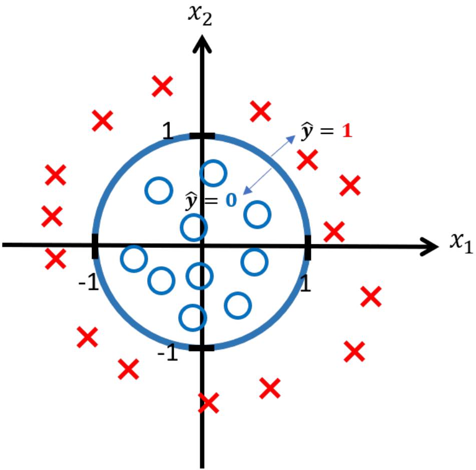
$$f_{\vec{w},b}(\vec{x}) = g(w_1 x_1^2 + w_2 x_2^2 + b)$$

Supondo $w_1 = w_2 = 1$ e $b = -1$, temos a seguinte fronteira de decisão:

$$z = \vec{w} \cdot \vec{x} + b = 0 \quad \rightarrow \quad x_1^2 + x_2^2 - 1 = 0 \quad \rightarrow \quad x_1^2 + x_2^2 = 1$$

Podemos ter também Fronteiras de Decisão mais complexas (não lineares):

Ilustrando a fronteira de decisão $x_1^2 + x_2^2 = 1$ na figura, temos



Podemos ter também Fronteiras de Decisão mais complexas (não lineares):

Observação

- Usando polinômios de maior ordem, podemos ter fronteiras de decisão com formas ainda mais complexas.
- Ou seja, assim como na Regressão Linear nós não estávamos limitados a estimar retar para os dados, aqui na Regressão Logística também não estamos limitados a estimar Fronteiras de Decisão lineares.
- Apesar disso, continua valendo a relação linear

$$z = \overrightarrow{w} \cdot \overrightarrow{x} + b$$

Let's say you are creating a tumor detection algorithm. Your algorithm will be used to flag potential tumors for future inspection by a specialist. What value should you use for a threshold?

- ○ High, say a threshold of 0.9?
- ◉ Low, say a threshold of 0.2?

Fonte: **Machine Learning Specialization**, *deeplearning.ai*, Stanford Online, Coursera.org.

Função Custo para Regressão Logística

Função Custo para Regressão Logística

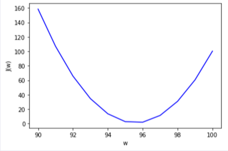
Na **Regressão linear**, tínhamos a seguinte função custo:

$$J(\vec{w}, b) = \frac{1}{2m} \sum_{i=1}^{m} \left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right) - y^{(i)} \right)^2$$

onde

$$f_{\vec{w},b}(\vec{x}) = \vec{w} \cdot \vec{x} + b$$

Vimos que, com essas definições, a função quadrática J é **convexa**, ou seja, ela não possui outros mínimos locais além do próprio mínimo global:

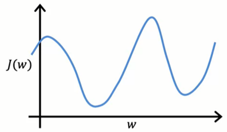


Agora, na **Regressão Logística**, temos:

$$f_{\vec{w},b}(\vec{x}) = \frac{1}{1 + e^{-(\vec{w} \cdot \vec{x} + b)}}$$

Pergunta: Agora para esse novo $f_{\vec{w},b}(\vec{x})$, é uma boa ideia usarmos a mesma definição para $J(\vec{w}, b)$?

Resposta: Não, afinal, é possível provar que, se fizermos isso, $J(\overrightarrow{w}, b)$ não será **convexa**, e terá diversos mínimos locais, aos quais podemos ficar presos.



Conclusão:

Precisaremos modificar um pouco nossa função custo $J(\overrightarrow{w}, b)$ para que ela se torne convexa agora com

$$f_{\overrightarrow{w},b}(\overrightarrow{x}) = \frac{1}{1 + e^{-(\overrightarrow{w} \cdot \overrightarrow{x} + b)}}$$

Função Custo para Regressão Logística

A função custo quadrática da **Regressão Linear** pode ser reescrita conforme segue:

$$J(\vec{w}, b) = \frac{1}{m} \sum_{i=1}^{m} L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right)$$

onde $L(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)})$ é chamada de "**função de perda**" e é dada por

$$L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right) = \frac{1}{2}\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right) - y^{(i)}\right)^2$$

Na **Regressão Logística**, também usamos a função custo:

$$J(\vec{w}, b) = \frac{1}{m} \sum_{i=1}^{m} L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right)$$

Porém, na Regressão Logística, a "**função de perda**" é dada por

$$L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right) = \begin{cases} -\log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) & \text{, se } y^{(i)} = 1 \\ -\log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) & \text{, se } y^{(i)} = 0 \end{cases}$$

OBS: Tal função é também chamada de "**função de entropia cruzada binária**".

Observação adicional: Estamos usando a notação $\log = \ln$ (logaritmo neperiano)

Analisando melhor a função perda

Na Regressão Logística, a **função de perda** é dada por

$$L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right) = \begin{cases} -\log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) & \text{, se } y^{(i)} = 1 \\ -\log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) & \text{, se } y^{(i)} = 0 \end{cases}$$

Observações:

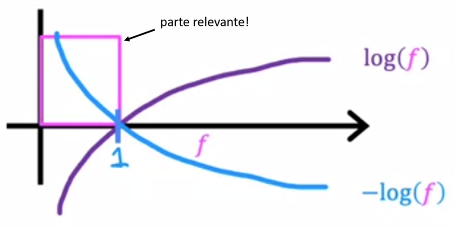
- a função de perda $L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right)$ opera com base em uma única amostra i
- Para obter a função custo, é necessário somar a perda para todas as amostras e depois dividir por m

Na Regressão Logística, a **função de perda** é dada por

$$L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right) = \begin{cases} -\log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) & \text{, se } y^{(i)} = 1 \\ -\log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) & \text{, se } y^{(i)} = 0 \end{cases}$$

Observações:

Como nossa $f_{\vec{w},b}\left(\vec{x}^{(i)}\right)$ estará sempre entre 0 e 1, então a única parte relevante da função $-\log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right)$ para nós é:



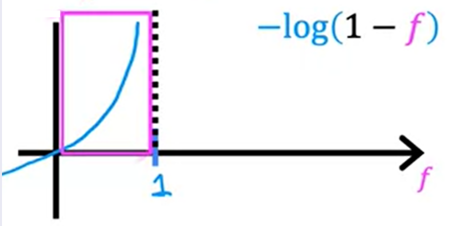
- Quando $y^{(i)} = 1$, a **perda** $\to 0$ conforme $f_{\vec{w},b}\left(\vec{x}^{(i)}\right) \to 1$. Ou seja, à medida com que a previsão torna-se correta! Por outro lado, **perda** $\to \infty$ conforme $f_{\vec{w},b}\left(\vec{x}^{(i)}\right) \to 0$.

Na Regressão Logística, a **função de perda** é dada por

$$L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right) = \begin{cases} -\log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) & \text{, se } y^{(i)} = 1 \\ -\log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) & \text{, se } y^{(i)} = 0 \end{cases}$$

Observações:

Como nossa $f_{\vec{w},b}\left(\vec{x}^{(i)}\right)$ estará sempre entre 0 e 1, então a única parte relevante da função $-\log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right)$ para nós é:



- Quando $y^{(i)} = 0$, a **perda** $\to 0$ conforme $f_{\vec{w},b}\left(\vec{x}^{(i)}\right) \to 0$. Ou seja, à medida com que a previsão torna-se correta! Por outro lado, **perda** $\to \infty$ conforme $f_{\vec{w},b}\left(\vec{x}^{(i)}\right) \to 1$.

Na Regressão Logística, a **função de perda** é dada por

$$L\left(f_{\overrightarrow{w},b}\left(\overrightarrow{x}^{(i)}\right), y^{(i)}\right) = \begin{cases} -\log\left(f_{\overrightarrow{w},b}\left(\overrightarrow{x}^{(i)}\right)\right) & \text{, se } y^{(i)} = 1 \\ -\log\left(1 - f_{\overrightarrow{w},b}\left(\overrightarrow{x}^{(i)}\right)\right) & \text{, se } y^{(i)} = 0 \end{cases}$$

- Definindo a função perda dessa forma, o modelo sofre uma penalidade zero (nula) quando sua previsão $f_{\overrightarrow{w},b}\left(\overrightarrow{x}^{(i)}\right)$ para uma determinada amostra $y^{(i)}$ está correta.
- A penalidade sofrida pelo modelo aumenta conforme sua previsão $f_{\overrightarrow{w},b}\left(\overrightarrow{x}^{(i)}\right)$ se afasta do valor alvo $y^{(i)}$
- Ao errar completamente a previsão, por exemplo, $y^{(i)} = 1$ e $f_{\overrightarrow{w},b}\left(\overrightarrow{x}^{(i)}\right) = 0$, o modelo é drasticamente penalizado (perda $\to \infty$)
- A função custo é a soma das perdas para todas as m amostras, divididos por m:

$$J(\overrightarrow{w}, b) = \frac{1}{m}\sum_{i=1}^{m} L\left(f_{\overrightarrow{w},b}\left(\overrightarrow{x}^{(i)}\right), y^{(i)}\right)$$

- O objetivo é encontrar os valores de $\overrightarrow{w}, b$ que minimizam essa perda média, $J(\overrightarrow{w}, b)$.

De olho no código!

Vamos agora explorar o conceito de Fronteira de Decisão e Função Custo para Regressão Logística.

Nome do arquivo que trabalharemos agora:

codigo - Fronteira de Decisão.ipynb

Parte 1

Rode todo o "codigo - Fronteira de Decisão.ipynb" sem fazer qualquer tipo de alteração. Certifique-se de que você o compreendeu.

Parte 2

1. Explique, com as suas próprias palavras, o conceito de Fronteira de Decisão no contexto da Regressão Logística.
2. Considerando $w_0 = w_1 = 1$ e $b = -3$, calcule o valor do modelo $f_{\overrightarrow{w},b}\left(\overrightarrow{x}^{(i)}\right)$ para cada amostra de dados $\overrightarrow{x}^{(i)}$ presente no código. O que esses valores representam? Os resultados estão coerentes com aquilo que é observado graficamente no código?
3. Calcule o valor da função perda para cada amostra i.
4. Calcule o custo $J(\overrightarrow{w}, b)$.