Regressão Logística: Função Custo Simplificada e Método do Gradiente

UFPR
UNIVERSIDADE FEDERAL DO PARANÁ

Engenharia Elétrica

Onde estamos e para onde vamos

Na aula anterior, definimos a **função perda** e a **função custo** para a Regressão Logística.

Nesta aula, iremos simplificar a definição da função custo e aplicar o Método do Gradiente para encontrar seu mínimo global.

Na aula passada, vimos que a **Função de perda** para Regressão Logística é dada por

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

OBS: $y^{(i)}$ é sempre 1 ou 0 (tumor maligno ou não)

Na aula passada, vimos que a **Função de perda** para Regressão Logística é dada por

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

OBS: $y^{(i)}$ é sempre 1 ou 0 (tumor maligno ou não)

Note que é possível simplificar, escrevendo da seguinte forma equivalente:

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

Pergunta:

Como a função de perda abaixo é simplificada para o caso em que a amostra i correspondente possui valor alvo $y^{(i)} = 1$?

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

A) $-\log\left(f_{\overrightarrow{w},b}\left(\overrightarrow{x}^{(i)}\right)\right)$

B) $-\log\left(f_{\overrightarrow{w},b}\left(1 - \overrightarrow{x}^{(i)}\right)\right)$

Função custo simplificada

A função custo que queremos minimizar na **Regressão Logística** é a média das perdas:

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

onde

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

Sendo assim, note que podemos reescrever

$$J(\overrightarrow{w}, b) = -\frac{1}{m} \sum_{i=1}^{m} \left[y^{(i)} \log\left(f_{\overrightarrow{w},b}\left(\overrightarrow{x}^{(i)}\right)\right) + (1 - y^{(i)}) \log\left(1 - f_{\overrightarrow{w},b}\left(\overrightarrow{x}^{(i)}\right)\right)\right]$$

Função custo simplificada

A função custo que queremos minimizar na Regressão Logística é a média das perdas:

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

onde

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

Sendo assim, note que podemos reescrever

$$J(\overrightarrow{w}, b) = -\frac{1}{m} \sum_{i=1}^{m} \left[y^{(i)} \log\left(f_{\overrightarrow{w},b}\left(\overrightarrow{x}^{(i)}\right)\right) + (1 - y^{(i)}) \log\left(1 - f_{\overrightarrow{w},b}\left(\overrightarrow{x}^{(i)}\right)\right)\right]$$

Observação:

- O Método do Gradiente agora pode ser aplicado para encontrar quais são os parâmetros $\overrightarrow{w}, b$ que minimizam $J(\overrightarrow{w}, b)$
- Lembrando que $J(\overrightarrow{w}, b)$ será convexa (único mínimo), ainda que o modelo seja:

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

onde $f_{\overrightarrow{w},b}(\overrightarrow{x})$ é a probabilidade de y ser 1.

Precisamos encontrar os valores de $\overrightarrow{w}, b$ que minimizam

$$J(\overrightarrow{w}, b) = -\frac{1}{m} \sum_{i=1}^{m} \left[y^{(i)} \log \left(f_{\overrightarrow{w}, b} \left(\overrightarrow{x}^{(i)} \right) \right) + (1 - y^{(i)}) \log \left(1 - f_{\overrightarrow{w}, b} \left(\overrightarrow{x}^{(i)} \right) \right) \right]$$

Precisamos encontrar os valores de $\overrightarrow{w}, b$ que minimizam

$$J(\overrightarrow{w}, b) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(f_{\overrightarrow{w},b}\left(\overrightarrow{x}^{(i)}\right)\right) + (1-y^{(i)})\log\left(1 - f_{\overrightarrow{w},b}\left(\overrightarrow{x}^{(i)}\right)\right)\right]$$

Sabemos que o Método do Gradiente consiste em repetir:

$$w_j = w_j - \alpha\frac{\partial}{\partial w_j}J(\overrightarrow{w}, b) \qquad j = 1, \cdots, n$$

$$b = b - \alpha\frac{\partial}{\partial b}J(\overrightarrow{w}, b)$$

Método do Gradiente aplicado à Regressão Logística

Precisamos encontrar os valores de $\overrightarrow{w}, b$ que minimizam

$$J(\overrightarrow{w}, b) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(f_{\overrightarrow{w},b}\left(\overrightarrow{x}^{(i)}\right)\right) + (1 - y^{(i)})\log\left(1 - f_{\overrightarrow{w},b}\left(\overrightarrow{x}^{(i)}\right)\right)\right]$$

Sabemos que o Método do Gradiente consiste em repetir:

$$w_j = w_j - \alpha\frac{\partial}{\partial w_j}J(\overrightarrow{w}, b) \qquad j = 1, \cdots, n$$

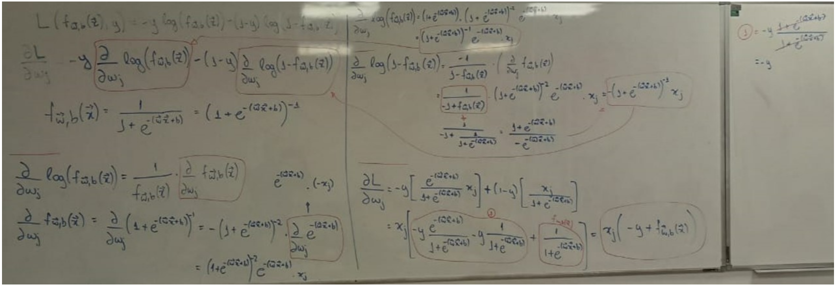
$$b = b - \alpha\frac{\partial}{\partial b}J(\overrightarrow{w}, b)$$

É possível provar que, para a função $J(\overrightarrow{w}, b)$ acima, tem-se:

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

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

Prova


$L(f_{\vec{w},b}(\vec{x}), y) = -y \log(f_{\vec{w},b}(\vec{x})) - (1-y) \log(1 - f_{\vec{w},b}(\vec{x}))$
$f_{\vec{w},b}(\vec{x}) = \frac{1}{1+e^{-(\vec{w}\cdot\vec{x}+b)}} = (1+e^{-(\vec{w}\cdot\vec{x}+b)})^{-1}$
$\frac{\partial L}{\partial w_j} = -y\left[\frac{e^{-(\vec{w}\cdot\vec{x}+b)}}{1+e^{-(\vec{w}\cdot\vec{x}+b)}} x_j\right] + (1-y)\left[\frac{x_j}{1+e^{-(\vec{w}\cdot\vec{x}+b)}}\right]$
$= x_j(-y + f_{\vec{w},b}(\vec{x}))$

Método do Gradiente para Regressão Logística consiste em repetir:

$$w_j = w_j - \alpha \left[\frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right) - y^{(i)} \right) x_j^{(i)} \right]$$

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

Parece idêntico à Regressão Linear. Porém, devemos lembrar que:

Regressão linear:

$$f_{\vec{w},b}\left(\vec{x}^{(i)}\right) = \vec{w} \cdot \vec{x} + b$$

Regressão Logística:

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

Quando estudamos Regressão Linear, é recomendado:

- Monitorar o Método do Gradiente (curva de aprendizado)
- Realizar implementação Vetorizada
- Escalonamento de Características

Todos estes conceitos continuam valendo agora para o contexto de Regressão Logística

De olho no código!

De olho no código!

Vamos agora implementar o algoritmo de Regressão Logística na prática.

Acesse o Python Notebook usando o QR code ou o link abaixo:

https://colab.research.google.com/github/xaximpvp2/master/blob/main/codigo_aula_11_regressao_logistica.ipynb

Atividade de aula

Parte 1

Rode todo o código. Certifique-se de que você o compreendeu.

Parte 2

1. Explique, com as suas próprias palavras, quais são os passos necessários para implementar a Regressão Logística na prática.
2. Faça modificações no conjunto de dados que está no código, e verifique como essas modificações alteram a fronteira de decisão do modelo estimado.