Problemas de Classificação

- Spam de email
- Transações fraudulentas / não fraudulentas.
- Diagnósticos tumor maligno \ benigno etc

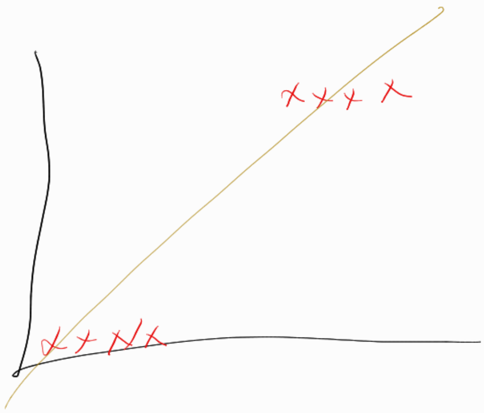
São exemplos binários, isto é:

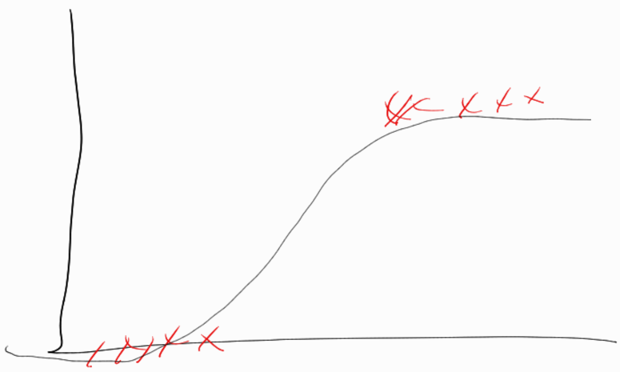
$$y \in \{0,1\}$$

ou generalizado

$$y \in \{0,1,2,3\ldots\}$$

Onde o problema base é:





Sendo necessário desenvolver um classificador $h(\theta,x)$

como ...

$$\begin{cases} h(\theta, x) \geq 0{,}5 \rightarrow y = 1 \\ h(\theta, x) < 0{,}5 \rightarrow y = 0 \end{cases}$$

Podem então usar a função sigmoide, que contém a característica necessária

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

Deste modo temos nossa hipótese escrita na forma:

$$h(\theta, x) = g(\theta^T x)$$

de modo $0 \leq h(\theta, x) \leq 1$

Em outras palavras

A probabilidade de $y = 1$ dado x parametrizado por θ

$$h(\theta, x) = P(y = 1 | x, \theta)$$

Como a probabilidade sempre assume valores entre 0 e 1, podemos normalizar de tal forma:

$$P(y=1|x,\theta) + P(y=0|x,\theta) = 1$$

$$P(y=0|x,\theta) = 1 - P(y=1|x,\theta)$$

CONSIDERANDO AGORA UM CONJUNTO DE DADOS

$$\{(x^{(1)}, y^{(1)}), x^{(2)}, y^{(2)} \ldots\}$$

$$x^{(i)} = \begin{pmatrix} x_0^{(i)} \\ x_1^{(i)} \\ x_2^{(i)} \\ \vdots \end{pmatrix}; \quad x_0^{(i)} = 1$$

NOSSA FUNÇÃO HIPÓTESE SENDO

$$h(\theta, x) = \frac{1}{1 + e^{-\theta^T}}$$

PODEMOS ATUALIZAR NOVAMENTE A FUNÇÃO CUSTO

$$J(\theta) = \frac{1}{m} \sum L\left(h(\theta, x^{(i)}), y^{(i)}\right)$$

SENDO QUE A FUNÇÃO CUSTO PARA A REGRESSÃO LINEAR NÃO FAZ MAIS SENTIDO, SENDO NECESSÁRIO MODIFICÁ-LA.

Para construla podemos primeiramente considerar

$$P(y=1 \mid x,\theta) = h(\theta,x) = \frac{1}{1+e^{-\theta^T x}}$$

$$P(y=0 \mid x,\theta) = 1-h(\theta,x) = \frac{1}{1+e^{+\theta^T x}}$$

usando distribuição de Bernoulli

$$P(y \mid x,\theta) = h^y(1-h)^{(1-y)}$$

Queremos ao fim da conta usar gradiente descendente para atualizar o theta, minimizando a função custo.

Para calcularmos então $P(\mathbf{y} \mid \mathbf{x},\theta)$, dado um conjunto $\mathbf{y}$ dado um conjunto $\mathbf{x}$ ajustado θ, sendo

$$P(\mathbf{y} \mid \mathbf{x},\theta) = P(y^{(1)} \mid x^{(1)},\theta)\, P(y^{(2)} \mid x^{(2)},\theta) \ldots P(y^{(m)} \mid x^{(m)},\theta)$$

escrito com outra notação

$$P(\mathbf{y} \mid \mathbf{x},\theta) = \prod_{i=0}^{m} P(y^{(i)} \mid x^{(i)},\theta)$$

sabendo que

$$J(\theta) = \frac{1}{m} \sum_{i=1} \mathcal{L}\left(h(\theta,x^{(i)}), y^{(i)}\right)$$

$J(\theta)$ já escrito em termos de um somatório, podemos aproveitar a propriedade do log

$$\log(a) + \log(b) = \log(ab)$$

Logo

$$J(\theta) = -\frac{1}{m}\log(P(y|X,\theta)) = -\frac{1}{m}\log\left(\prod P(y|X,\theta)\right)$$

$$= -\frac{1}{m}\sum \log\left(P(y|X,\theta)\right) = -\frac{1}{m}\sum \log\left(h^y(1-h)^{(1-y)}\right)$$

$$= -\frac{1}{m}\sum\left[\log\left(h^y\right) + \log\left(1-h^{(1-y)}\right)\right]$$

$$= -\frac{1}{m}\sum\left[y\log(h) + (1-y)\log(1-h)\right]$$

Podemos finalmente aplicar o GD derivando a função custo:

$$\theta_j = \theta_j - \eta\frac{\partial J}{\partial \theta_j}$$

$$\frac{\partial J}{\partial \theta_j} = -\frac{1}{m}\sum\frac{\partial}{\partial \theta_j}\left[y\log(h) + (1-y)\log(1-h)\right]$$

$\hookrightarrow h(\theta)$

COMO h DEPENDE DE θ

$$= -\frac{1}{m}\sum \frac{\partial h}{\partial \theta}\frac{\partial}{\partial h}\left[\cdots\right]$$

$$= -\frac{1}{m}\sum \frac{\partial h}{\partial \theta}\left[y\frac{1}{h} - \frac{(1-y)}{1-h}\right]$$

$$= -\frac{1}{m}\sum \frac{\partial h}{\partial \theta}\left[\frac{y-\cancel{hy}-h+\cancel{hy}}{h(1-h)}\right]$$

$$= \frac{1}{m}\sum \frac{\partial}{\partial \theta}\left(\frac{1}{1+e^{-\theta^T x}}\right)\left[\frac{h-y}{h(1-h)}\right]$$

$$= \frac{1}{m}\sum \frac{0\cdot\cancel{(1+e^{-\theta^T x})} + x e^{-\theta^T x}}{(1+e^{-\theta^T x})^2}$$

$$= \frac{1}{m}\sum \underbrace{\left(-x e^{-\theta^T x}\right)\left(\frac{1}{1+e^{-\theta^T x}}\right)^2}_{\frac{1}{e^{x}+1}}\left[\frac{h-y}{\cancel{h}(1-h)}\right]$$

SABENDO QUE $1-h = \dfrac{1}{1+e^{\theta^T x}} \rightarrow 1+e^{\theta^T} = \dfrac{1}{(1-h)}$

$$h = \frac{1}{1+e^{-\theta^T x}}$$

$$\frac{\partial J}{\partial \theta} = \frac{1}{m}\sum\left(\frac{1}{\cancel{e^{\theta^T}+1}}\right)\left(\cancel{1+e^{\theta^T}}\right)\left[h-y\right]x$$

$$\boxed{\frac{\partial J}{\partial \Theta} = \frac{1}{m} \sum \left[h - y \right] x}$$