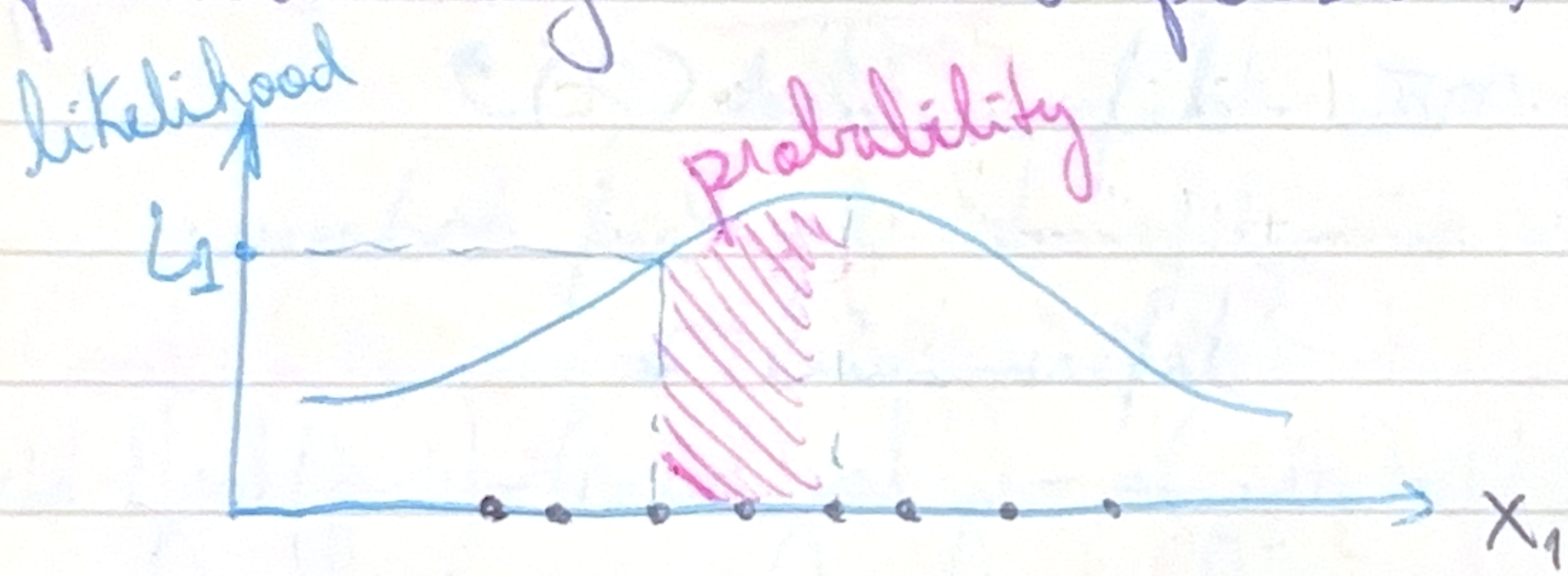
MLE: maximum likelihood estimation

↓

dado um conjunto de dados e um modelo estatístico, o MLE permite estimar os PARÂMETROS do modelo.

Basicamente o método permite estimar os parâmetros com os quais a probabilidade de se observar os dados disponíveis seja a maior possível.




$$\frac{\partial J(\theta)}{\partial \theta_j} = -\frac{1}{m}\sum_{i=1}^{m}\left(\frac{y^{(i)}\,h'_\theta(x^{(i)})}{h_\theta(x^{(i)})} - \frac{(1-y^{(i)})\,.\,h'_\theta(x^{(i)})}{1-h_\theta(x^{(i)})}\right)$$

e como $h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$, então $\frac{\partial h_\theta(x)}{\partial \theta_j} = \frac{+x_j^{(i)}\,.\,e^{-\theta^T x}}{(1+e^{-\theta^T x})^2}$

$$\frac{\partial J(\theta)}{\partial \theta_j} = -\frac{1}{m}\sum_{i=1}^{m}\frac{\left[y^{(i)} - \cancel{h_\theta(x)y^{(i)}} - h_\theta(x^{(i)}) + \cancel{h_\theta(x^{(i)})y^{(i)}}\right] h'_\theta(x^{(i)})}{h_\theta(x^{(i)})(1-h_\theta(x^{(i)}))}$$

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m}\frac{(h_\theta(x^{(i)}) - y^{(i)})\,.\,h'_\theta(x^{(i)})}{h_\theta(x^{(i)})(1-h_\theta(x^{(i)}))}$$

$$= \frac{1}{m}\sum_{i=1}^{m}\frac{(h_\theta(x^{(i)}) - y^{(i)})\,.}{\left(\frac{1}{1+e^{-\theta^T x}}\right)\cdot\left(1-\frac{1}{1+e^{-\theta^T x}}\right)}\,\frac{x_j^{(i)}\,.\,e^{-\theta^T x}}{(1+e^{-\theta^T x})^2}$$

$$= \frac{1}{m}\sum_{i=1}^{m}\frac{(h_\theta(x^{(i)}) - y^{(i)})}{\frac{\cancel{1}+e^{-\theta^T x}-\cancel{1}}{(1+e^{-\theta^T x})}}\,.\,\frac{x_j^{(i)}\,.\,e^{-\theta^T x}}{(1+e^{-\theta^T x})}$$

$$= \boxed{\frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)x_j^{(i)}}$$

Consegui!!!

$$\boxed{\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)x_j^{(i)}}$$

for GRADIENT DESCENT