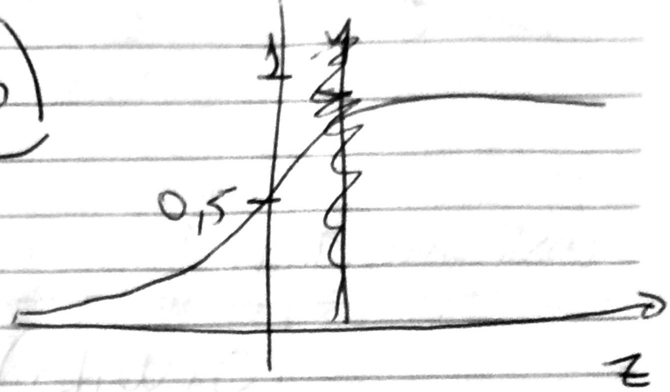
Aula 5 | Intuition About log reg

Given X, $\hat{y} = P(y=1 \mid x)$

$x \in \mathbb{R}^{n_x}$

Parameter: $w \in \mathbb{R}^{n_x}$, $b \in \mathbb{R}$

$\hat{y} = \sigma(\underbrace{w^T x + b}_{z})$



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

| $z \to \infty$, $\sigma(z) \to 1$ | $z \to -\infty$, $\sigma \to 0$.

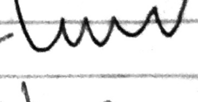
Cost function

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

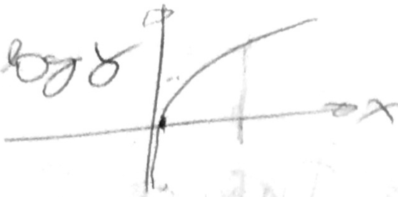
We want $\hat{y}^{(i)} \approx y^{(i)}$. At least in the training set.

In logistic regression: $L(\hat{y}, y) = \frac{1}{2}(\hat{y} - y)^2$

don't do that



local optima



$$L(\hat{y}, y) = -(y \log \hat{y} + (1-y)\log(1-\hat{y}))$$

If $y=1$: $-(1 . \log \hat{y})$

↳ large. close 1

If $y=0$: $-\log(1-\hat{y})$

↳ small

$\hat{y} \to 0$.

Cost function:

$$\boxed{J(w,b)} = \sum_{i=1}^{m} \frac{L(\hat{y}^{(i)}, y^{(i)})}{m}$$

$$= \frac{-1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\hat{y}^{(i)} + (1-y^{(i)})\log(1-\hat{y}^{(i)})\right]$$

Gradient Descent

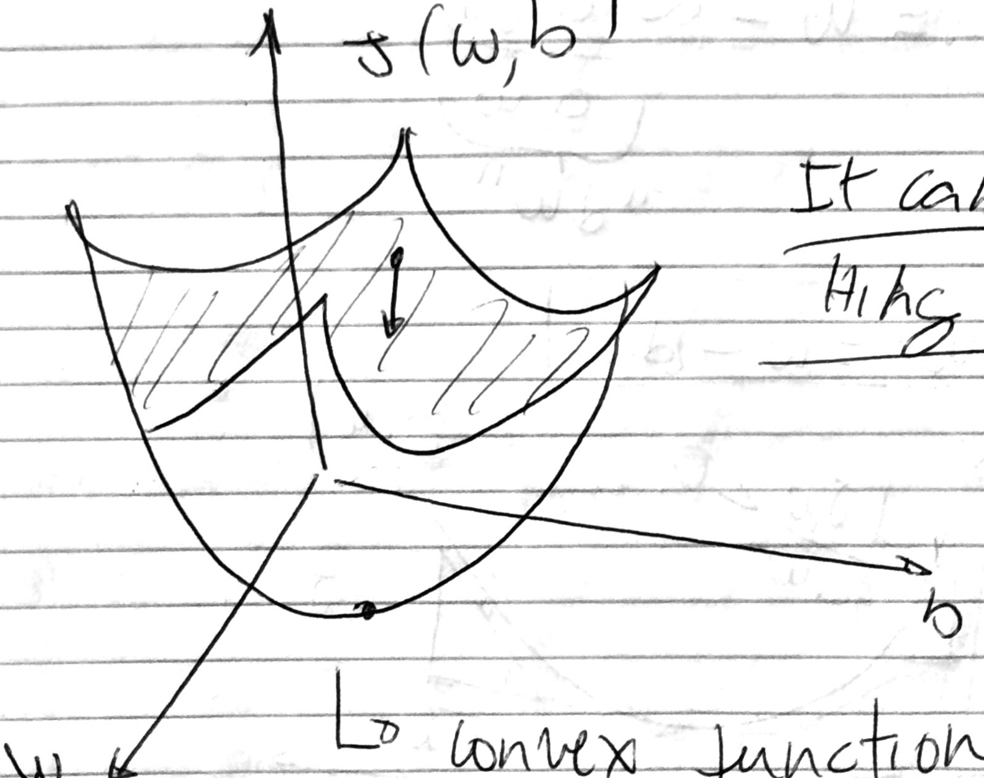
$\hat{y} = \sigma(w^T x + b)$

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

$$J(w,b) = \frac{1}{m}\sum_{i=1}^{m} L(\hat{y}^{(i)}, y^{(i)}) =$$

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

We want to find → w, b that minimize J

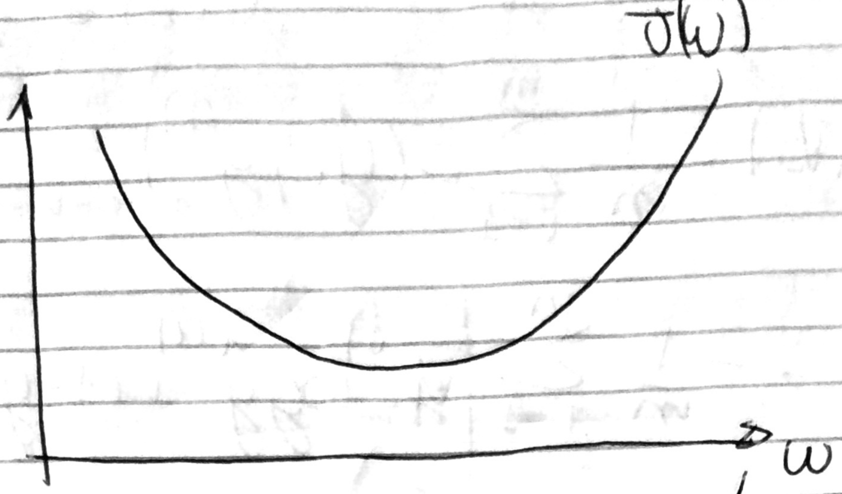


It can be high dimensional

↳ convex function

As the function is convex any pair (w, b) is acceptable.

Gradient desent



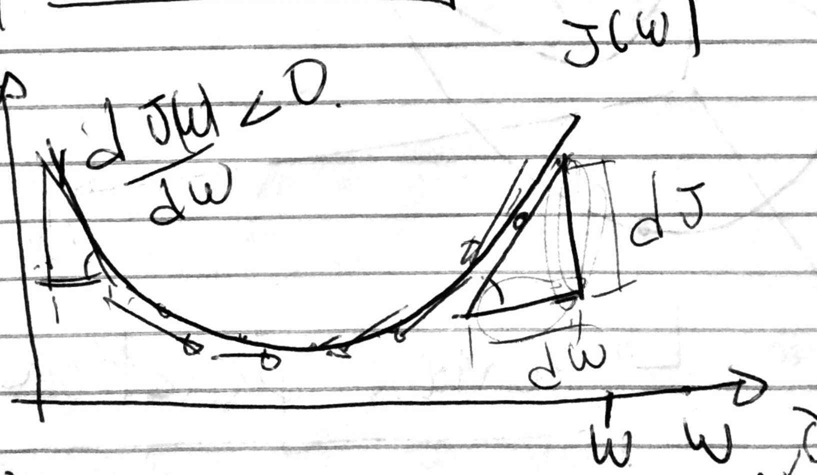
learnin rate

Repeat {

$$w := w - \alpha \underbrace{\frac{dJw}{dw}}_{\text{"dw"}}$$

}

$$\boxed{w := w - \alpha dw}$$

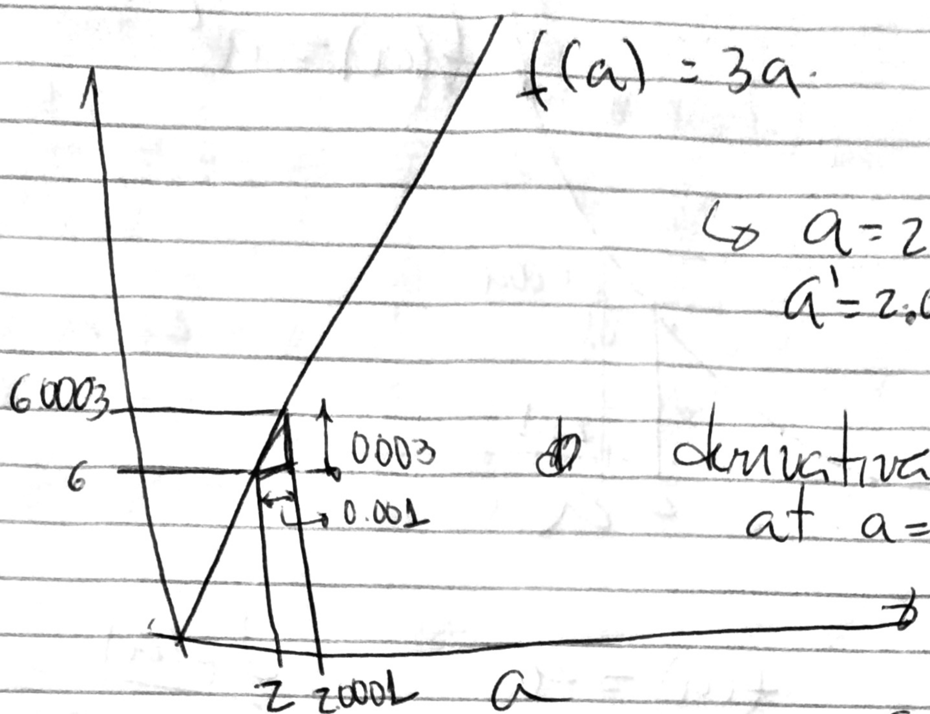


$$J(w,b) \Rightarrow \quad w := w - \alpha \frac{\partial J(w,b)}{\partial w}$$

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

derivadas

$f(a) = 3a$



↳ $a = 2$ $\quad f(a) = 6$

$a' = 2.0001$ $\quad f(a') = 6.0003$

derivativa of $f(a)$ at $a = 2$ $= \dfrac{0.0003}{0.0001} = 3$

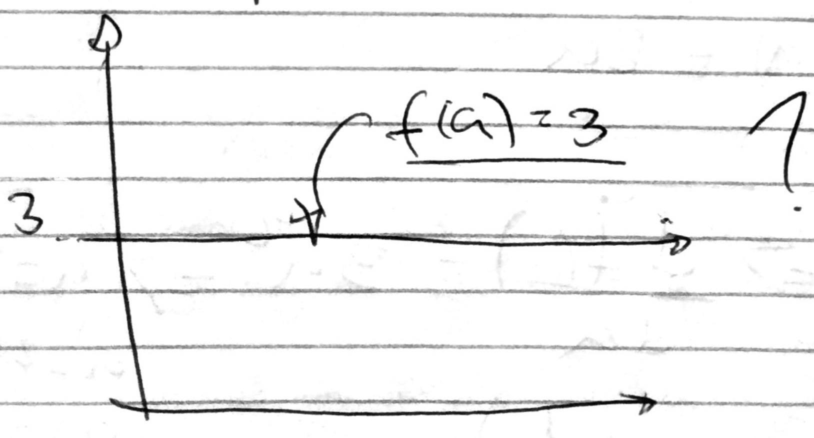
$a = 5$ $\quad f(a) = 15$

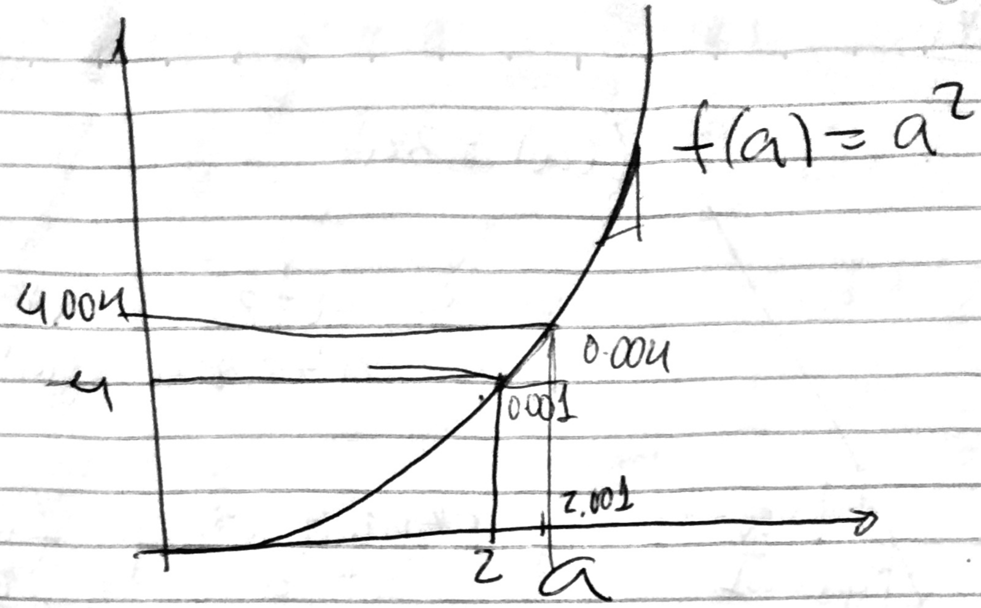
$a' = 5.0001$ $\quad f(a') = 15.0003$

slope = 3

$$\boxed{\frac{df(a)}{da} = 3}$$

↳ mover (a) pra direita um pouco f(a) aumenta 3x esse pouco.





$a = 2$ $\qquad f(a) = 4$ $\qquad \left.\dfrac{d f(a)}{d a}\right|_{a=4} = \dfrac{0.004}{0.001}$

$a' = 2.001$ $\qquad f(a) \cong 4.004$

$= \cancel{0}4$

$a = 5$ $\qquad f(a) = 25$ $\qquad \left.\dfrac{d f(a)}{d a}\right|_{a=5} = 10.$

$a = 5.001$ $\qquad f(a) \cong 25.0\underline{10}$

$$\frac{d f(a)}{d a} = 2a$$

$$f(a) = a^3 \Rightarrow \frac{d f(a)}{d a} = 3a^2 \Rightarrow a = 2$$

$a' = 2.001$

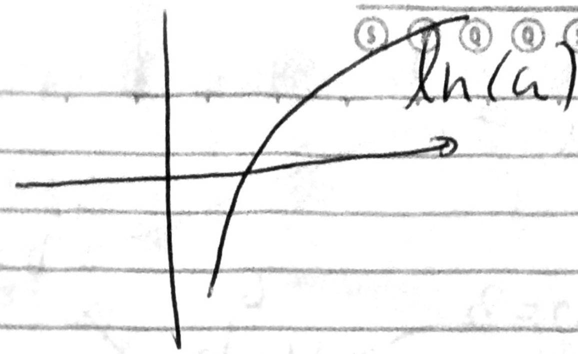
$$\left.\frac{d f(a)}{d a}\right|_{a=2} = 12$$

$f(a) = 8$

$f(a') = 8.\underline{012}$

$f(a) = \log_e(a)$

$\ln(a)$



$$\boxed{\frac{df(a)}{da} = \frac{1}{a}}$$

$a = 2 \quad f(a) = 0.69315$

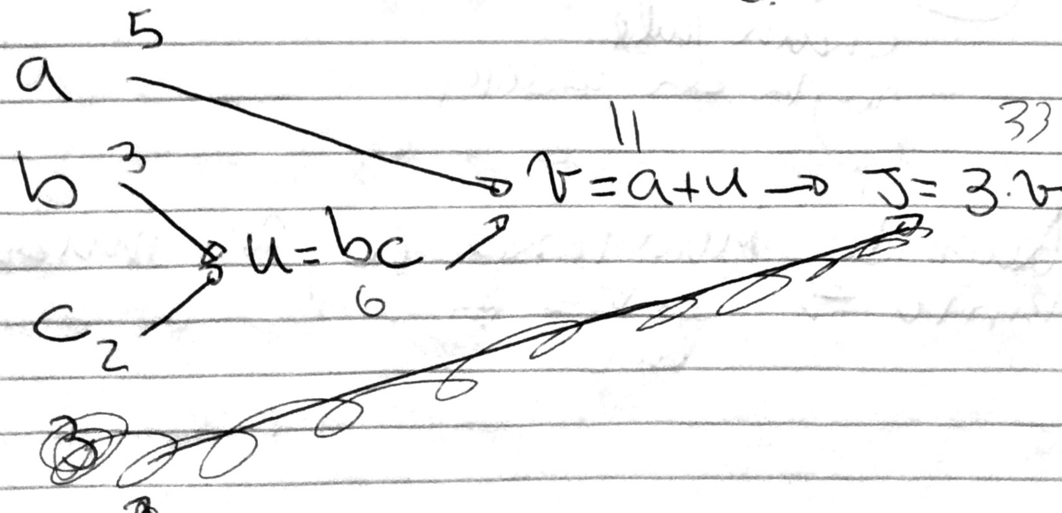
$a' = 2.001 \quad f(a') = 0.69365$

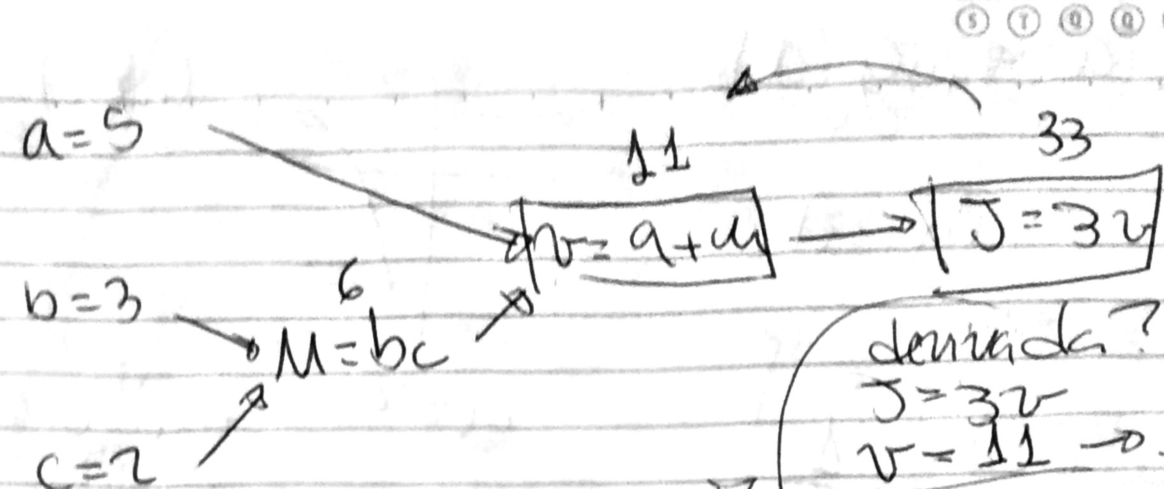
$\frac{1}{2} \rightarrow 0.0005$

Computation GRAPHS

$$J(a,b,c) = \underbrace{3(\underbrace{a + \underbrace{bc}_{u}}_{v})}_{J}$$

$u = bc$
$v = a + u$
$J = 3v.$





derivada?

$J=3v$

$v=11 \to 11.001$

$J=33 \to 33.003$

① $\frac{dJ}{dv}=?$ take derivative

② $\frac{dJ}{da}?$

$a=5 \to 5.001$

$v=11 \to 11.001$

$J=33 \to 33.003$

$\frac{dJ}{da}=3$

$$\underbrace{\frac{dJ}{da}=\frac{dJ}{dv}\cdot\frac{dv}{da}}_{\text{chain rule}}=\underline{3}$$

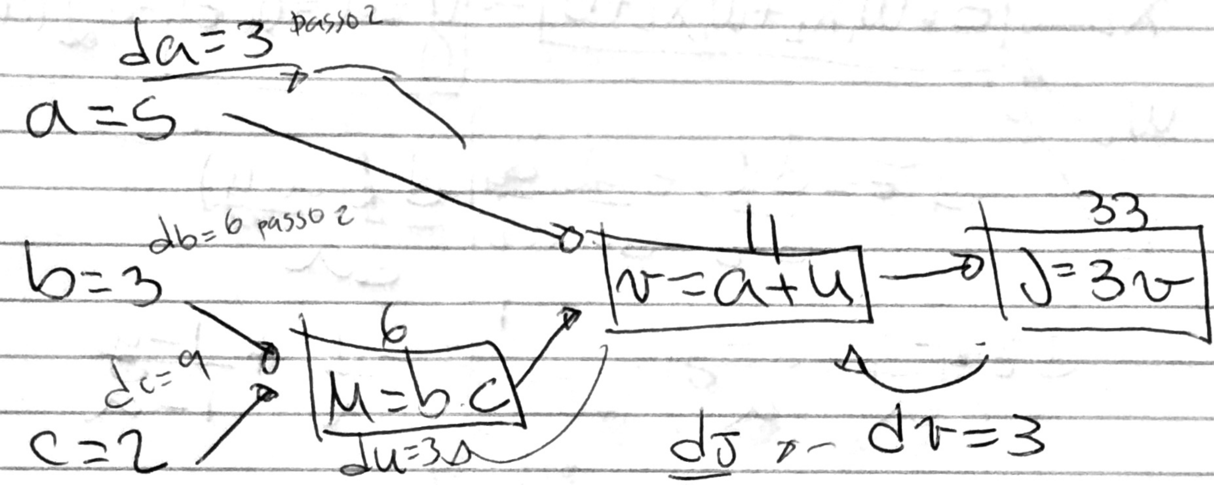
chain rule
regra da cadeia.

Quando Mudamos a, v muda o mesmo tanto $\Rightarrow \frac{dv}{da}=1$.

We care about a final output variable

$$\boxed{\frac{d\ \text{Final Output Variable}}{d\,var}}$$

notation dvar is the derivative in the code.



$\frac{dJ}{dv} \rightarrow dv=3$

$$\frac{dJ}{du} = ? = 3 = \underbrace{\frac{dJ}{dv}}_{3} \cdot \underbrace{\frac{dv}{du}}_{1}$$

$u = 6 \rightarrow 6.001$
$v = 11 \rightarrow 11.001$
$J = 33 \rightarrow 33.003$

regra da cadeia

~~du~~

$$\frac{dJ}{db} = \frac{dJ}{du} \cdot \underbrace{\frac{du}{db}}_{2} \Rightarrow$$

$b = 3 \rightarrow 3.001$
$u = b \cdot c = 6 \rightarrow 6.002$

$\frac{du}{db} = 2$

$$\frac{dJ}{dc} = \underbrace{\frac{dJ}{du}}_{3} \cdot \underbrace{\frac{du}{dc}}_{3} = 9.$$

Logistic Regression Gradient Descent

x_1, w_1, x_2, w_2, b $\rightarrow$ $\boxed{z = w_1x_1 + w_2x_2 + b}$ $\rightarrow$ $\boxed{\hat{y} = a = \sigma(z)}$ $\rightarrow$ $\boxed{\mathcal{L}(a,y)}$

$$dz = \frac{d\mathcal{L}}{dz} = \frac{d\mathcal{L}}{da}\cdot\frac{da}{dz} \qquad \frac{d\mathcal{L}(a,y)}{da}$$

$$dz = \left(\frac{-y}{a} + \frac{1-y}{a}\right)\cdot\left(\frac{1}{1+e^{-z}}\right)' = \frac{-y}{a} + \frac{1-y}{1-a}$$

$$\boxed{a(1-a)}$$

$$\Rightarrow a - y$$

$$\frac{d\mathcal{L}}{dw_1} = \text{"}dw_1\text{"} = \frac{d\mathcal{L}}{dz}\cdot\frac{dz}{dw_1} = (a-y)\cdot x_1$$

$$\frac{d\mathcal{L}}{dw_2} = \text{"}dw_2\text{"} = dz \cdot x_2 = (a-y)\cdot x_2$$

$$\boxed{db = dz}$$

$$w_1 := w_1 - \alpha\, dw_1$$
$$w_2 := w_2 - \alpha\, dw_2$$
$$\vdots$$
$$w_n := w_n - \alpha\, dw_n$$
$$\cancel{b := b - \alpha}$$
$$b = b - \alpha\, db$$

~~Regressão para m exemplos~~ Logística

$$J(w,b) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}(a^{(i)}, y^{(i)})$$ → média dos erros

direção para onde o erro está caindo

$$\frac{\partial J(w,b)}{\partial w_i} = \frac{1}{m}\sum_{i=1}^{m} \underbrace{\frac{\partial \mathcal{L}(a^{(i)}, y^{(i)})}{\partial w_i}}_{dw_i^{(i)} - (x^{(i)}, y^{(i)})}$$