$$\bar{y} = f(w^T x + b)$$

Loss error function

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

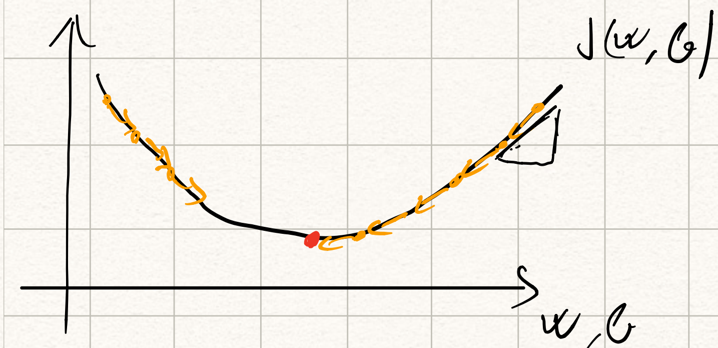
Cost function

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

↳ A predikció mennyire jól közelíti a ground truth értéket

↳ keressük azt a (w, b)-t ami minimalizálja $J(w,b)$-t

Gradient descent



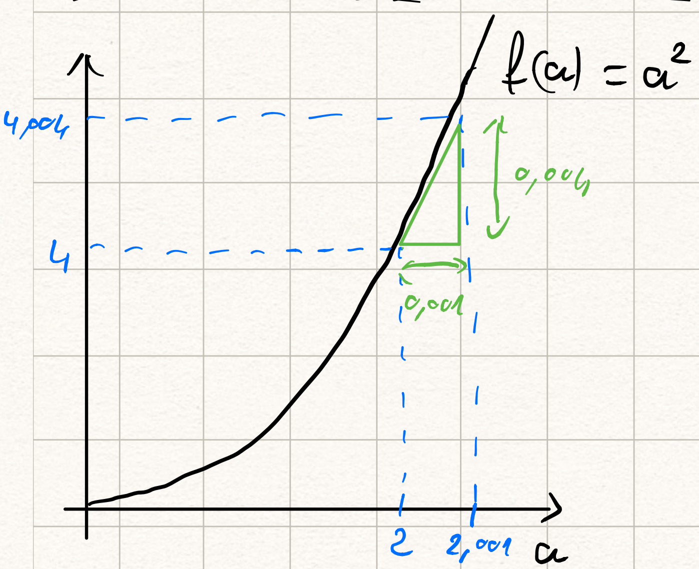
Repeat {

$$w := w - \alpha \frac{dJ(w,b)}{dw}$$

$$b := b - \alpha \frac{dJ(w,b)}{db}$$

}

Intuition about derivatives (slope of f(a))



1) $a = 2, \ f(a) = 4$

$$\frac{d}{da} f(a) = 4$$

2) $a = 5, \ f(a) = 25$

$$\frac{d}{da} f(a) = 10$$

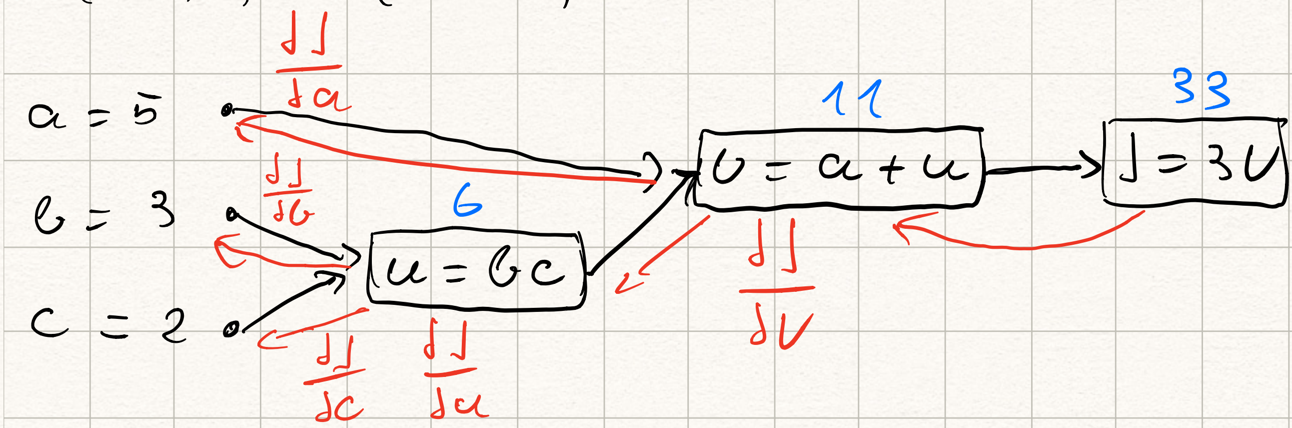
$$\frac{d}{da} f(a) = \frac{d}{da} a^2 = 2a$$

Computation graph

$\longrightarrow$ forward path

$\longleftarrow$ backward path

$$J(a,b,c) = 3(a+bc)$$



$J = 3v$

$v = 11 \rightarrow J = 33$ $\Rightarrow$ $\dfrac{dJ}{dv} = 3$

$v' = 11{,}001 \rightarrow J' = 33{,}003$

$a = 5 \rightarrow 5{,}001$

$v = 11 \rightarrow 11{,}001$ $\Rightarrow$ $\dfrac{dJ}{da} = \dfrac{dJ}{dv}\,\dfrac{dv}{da} = 3 \cdot 1 = 3$

$J = 33 \rightarrow 33{,}003$

chain - rule

$u = 6 \rightarrow 6{,}001$

$v = 11 \rightarrow 11{,}001$ $\Rightarrow$ $\dfrac{dJ}{du} = \dfrac{dJ}{dv}\,\dfrac{dv}{du} = 3 \cdot 1 = 3$

$J = 33 \rightarrow 33{,}003$

$b = 3 \rightarrow 3{,}001$

$c = 2$ $\Rightarrow$ $\dfrac{dJ}{db} = \dfrac{dJ}{du}\,\dfrac{du}{db} = 3 \cdot 2 = 6$

$u = bc = 6 \rightarrow 6{,}002$

$$\frac{dJ}{dc} = \frac{dJ}{du}\,\frac{du}{dc} = 3 \cdot 3 = 9$$

Logistic regression derivatives

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

$$dz = \frac{d\mathcal{L}(a,y)}{dz} = \frac{d\mathcal{L}(a,y)}{da} \underbrace{\frac{da}{dz}}_{a(1-a)} = a - y$$

$$da = \frac{d\mathcal{L}(a,y)}{da} = -\frac{y}{a} + \frac{1-y}{1-a}$$

$$\frac{d\mathcal{L}}{dw_1} = dw_1 = x_1\,dz$$

$$dw_2 = x_2\,dz$$

$$db = dz$$

$$\Rightarrow \quad w_1 := w_1 - \alpha\, dw_1$$

$$w_2 := w_2 - \alpha\, dw_2$$

$$b := b - \alpha\, db$$

Vectorization for speed

```
z = 0
for i in range(n):
    z += w[i] x x[i]
z += b
```

$$\Rightarrow \quad z = \underbrace{np.dot(w, x)}_{w^T x} + b$$