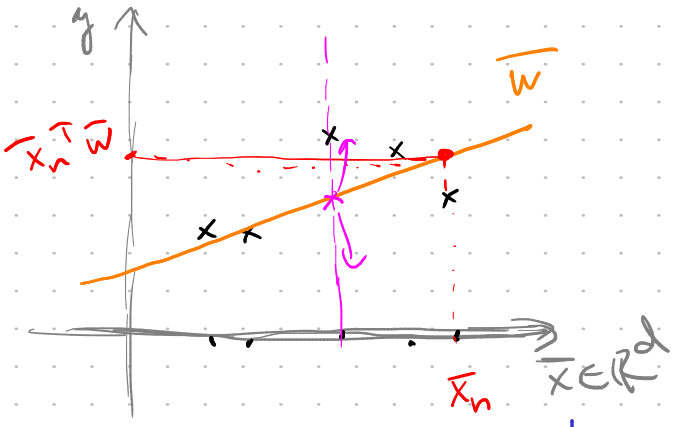
① Linear regression

$$D = \{(\bar{x}_n, y_n)\}_{n=1}^{N}$$



$$y \approx w_0 + w_1 x_1 + w_2 x_2 + \dots + w_d x_d = \bar{x}^T \bar{w}$$

$$\text{Error}(\bar{x}, y) = |y - \bar{w}^T \bar{x}|$$

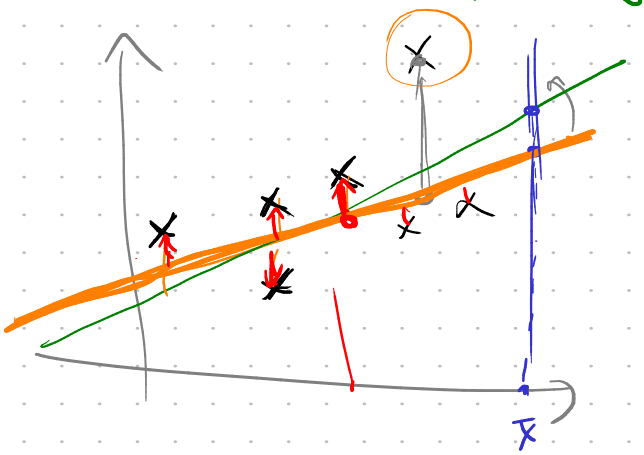
$$L(\bar{w}) = \sum_{n=1}^{N} |y_n - \bar{w}^T \bar{x}_n| \xrightarrow{\bar{w}} \min$$

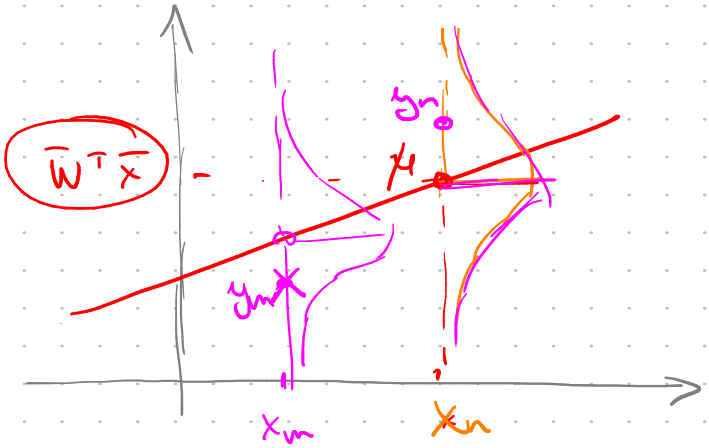
$f(\bar{w}, \bar{x}_n)$

$$L(\bar{w}) = \sum_n (y_n - \bar{w}^T \bar{x}_n)^2$$

$$L(\bar{w}) = \sum_n \left(e^{|y_n - \bar{w}^T \bar{x}_n|} - 1\right)$$

$$L(\bar{w}) = \sum_n (y_n - \bar{w}^T \bar{x}_n)^4$$





Likelihood: $p(D|\bar{w}) = \prod_{n=1}^{N} p(y_n | \bar{w}, \bar{x}_n) =$

$$N(z|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(z-\mu)^2}$$

$$= \prod_n N(y_n | \bar{w}^T \bar{x}_n, \sigma^2) = \prod_n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_n - \bar{w}^T \bar{x}_n)^2}$$

$$\log p(D|\bar{w}) \xrightarrow{\bar{w}} \max$$

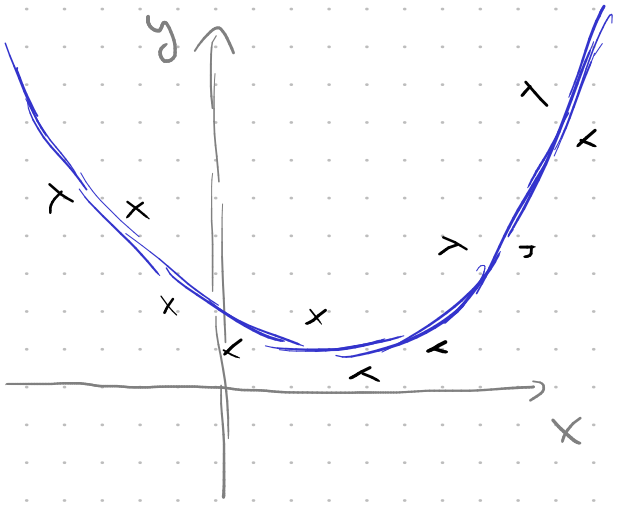
$$\log p(D|\bar{w}) = \sum_n \left(-\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y_n - \bar{w}^T \bar{x}_n)^2 \right) =$$

$\log p(D|\bar{w})$ → $\max_{\bar{w}}$

$$= \underbrace{-\frac{N}{2}\log 2\pi\sigma^2}_{\text{Const}} - \underbrace{\frac{1}{2\sigma^2}}_{\text{Const}} \boxed{\sum_{n=1}^{N}(y_n - \bar{w}^T \bar{x}_n)^2}$$

$\bar{w}$ → $\min$

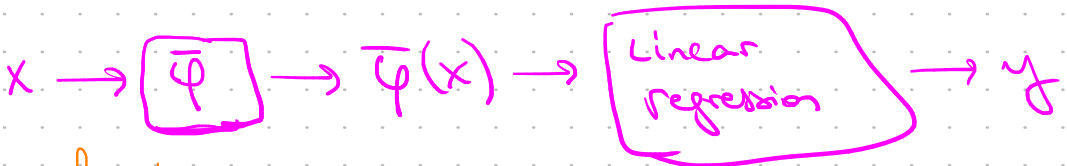
(2) Features, overfitting, regularization



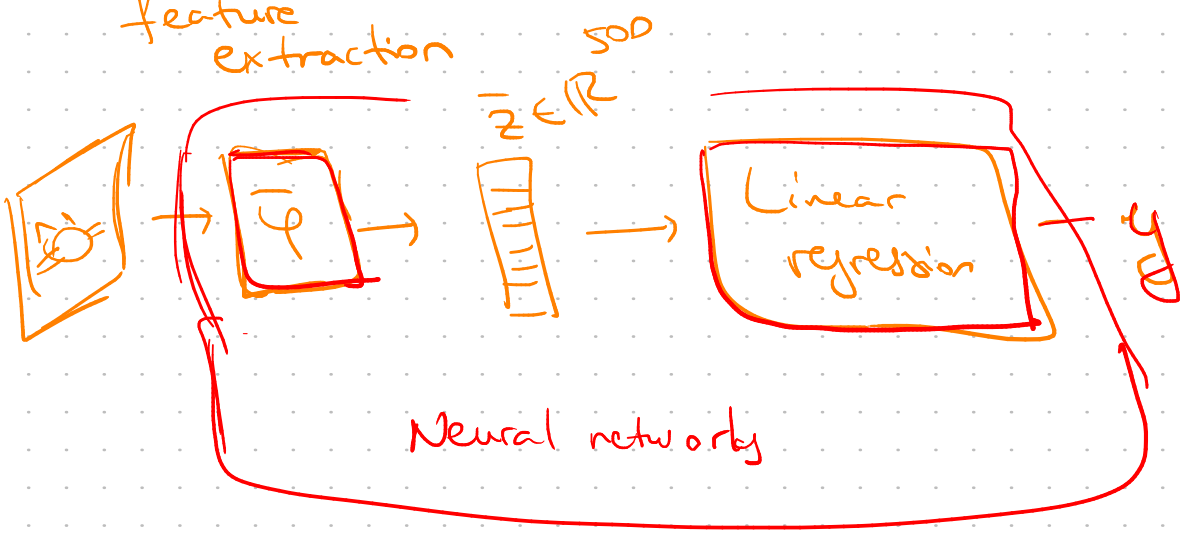
$$D = \{(x_n, y_n)\}_{n=1}^{N}$$

$$y \approx w_0 + w_1 x + w_2 x^2 = \begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix}^T \begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix}$$

$$x \in \mathbb{R} \rightsquigarrow \begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix} \in \mathbb{R}^3$$



feature extraction



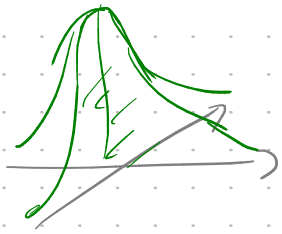
$x \in \mathbb{R} \rightsquigarrow \bar{\varphi} \rightsquigarrow \begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix} \in \mathbb{R}^3$

$\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_d \end{pmatrix} \in \mathbb{R}^d \quad (d = 100) \rightsquigarrow \bar{\varphi} \rightsquigarrow$

$\begin{pmatrix} 1 \\ x_1 \\ \dots \\ x_d \\ x_1^2 \\ x_1 x_2 \\ x_1 x_3 \\ \dots \\ x_d^2 \end{pmatrix}$ $\left.\begin{matrix} \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{matrix}\right\} d+1$ $\left.\begin{matrix} \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{matrix}\right\} \frac{d(d+1)}{2}$ $\approx \frac{d^2}{2}$

$T^{\max}$

$p(\bar{w}|D) = \dfrac{p(D|\bar{w})\, p(\bar{w})}{p(D)}$ — $p(D|\bar{w})$: likelihood (max), $p(\bar{w})$: prior, $p(\bar{w}|D)$: posterior



$\log p(\bar{w}|D) = \text{const} + \log p(D|\bar{w}) + \log p(\bar{w})$

$\log p(D|\bar{w})$: $\bar{w}$, max

$\log p(\bar{w})$: $\bar{w}$, max

$L(\bar{w}) = \sum_n (y_n - \bar{w}^T \bar{x}_n)^2 - \log p(\bar{w})$

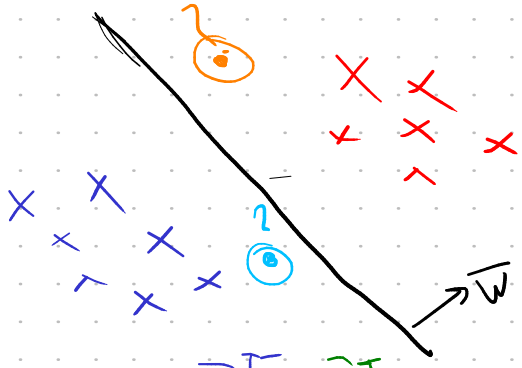
$\bar{w}$, min

regularizer

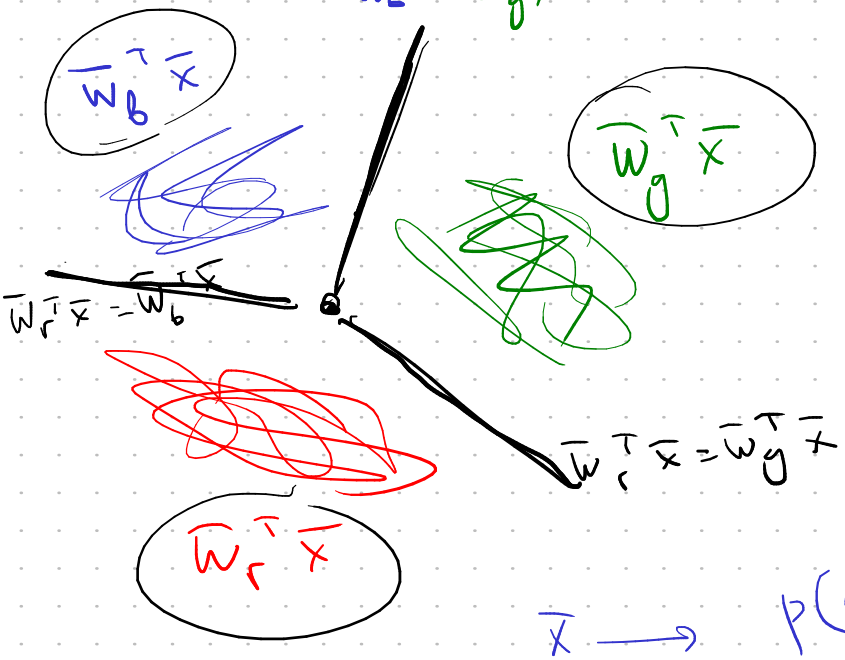
$L(\bar{w}) = \sum_n (y_n - \bar{w}^T \bar{x}_n)^2 + \alpha \sum_{i=1}^{d} w_i^2$ — L_2-reg

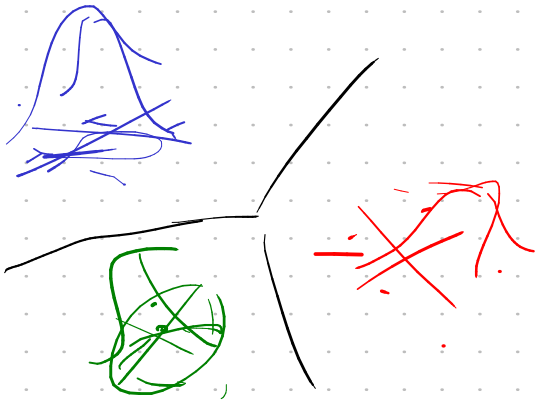
$+ \alpha \sum_{i=1}^{d} |w_i|$ — L_1-reg.

② Classification









$$\bar{x} \longrightarrow p(C_1|\bar{x}), p(C_2|\bar{x}), \ldots, p(C_k|\bar{x})$$

C_1, C_2

$$p(C_1|\bar{x}) = \frac{p(\bar{x}|C_1)\,p(C_1)}{p(\bar{x}|C_1)p(C_1) + p(\bar{x}|C_2)\,p(C_2)} =$$

$$= \frac{1}{1 + \dfrac{p(\bar{x}|C_2)p(C_2)}{p(\bar{x}|C_1)p(C_1)}} = \frac{1}{1 + e^{-\log \frac{p(\bar{x}|C_1)p(C_1)}{p(\bar{x}|C_2)p(C_2)}}} = \frac{1}{1 + e^{-\bar{w}^T \bar{x}}}$$

log-odds

$\dfrac{p(C_2|\bar{x})}{p(C_1|\bar{x})}$ — шансы, odds

$$p(C_1|\bar{x}) = \frac{1}{1+e^{-\bar{w}^T\bar{x}}} = \sigma(\bar{w}^T\bar{x})$$

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

$$p(C_2|\bar{x}) = 1 - \sigma(\bar{w}^T\bar{x})$$

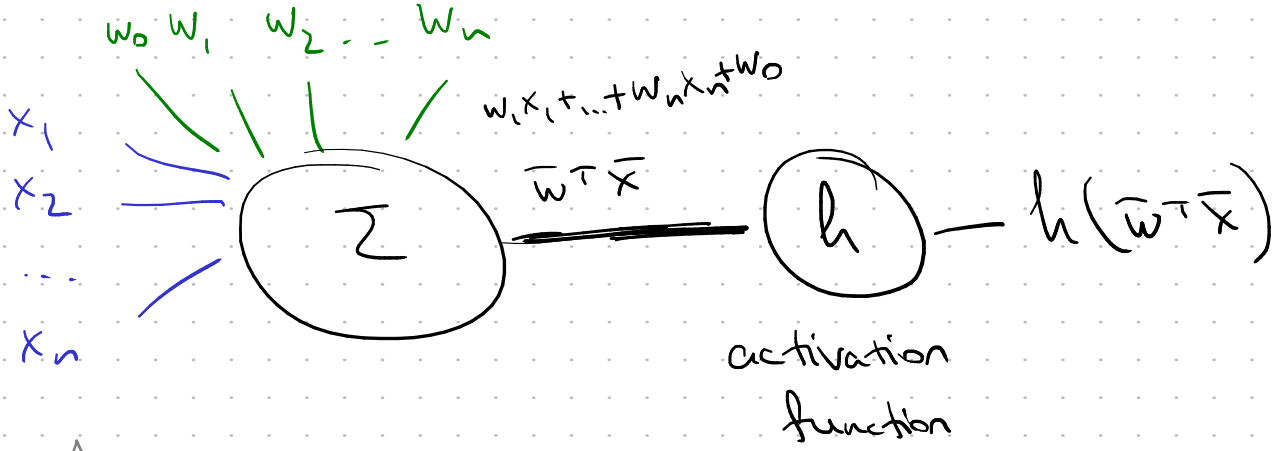
$$p(C_k|\bar{x}) = \frac{p(\bar{x}|C_k)p(C_k)}{\sum_s p(\bar{x}|C_s)p(C_s)} = \frac{e^{\log p(\bar{x}|C_k)p(C_k)}}{\sum_s e^{\log p(\bar{x}|C_s)p(C_s)}}$$

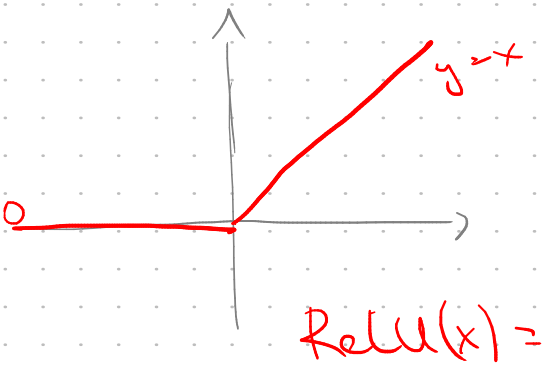
$\log p(\bar{x}|C_k)p(C_k) = \bar{w}_k^T\bar{x} \in (-\infty, \infty)$

$\log p(\bar{x}|C_s)p(C_s) \simeq \bar{w}_s^T\bar{x}$

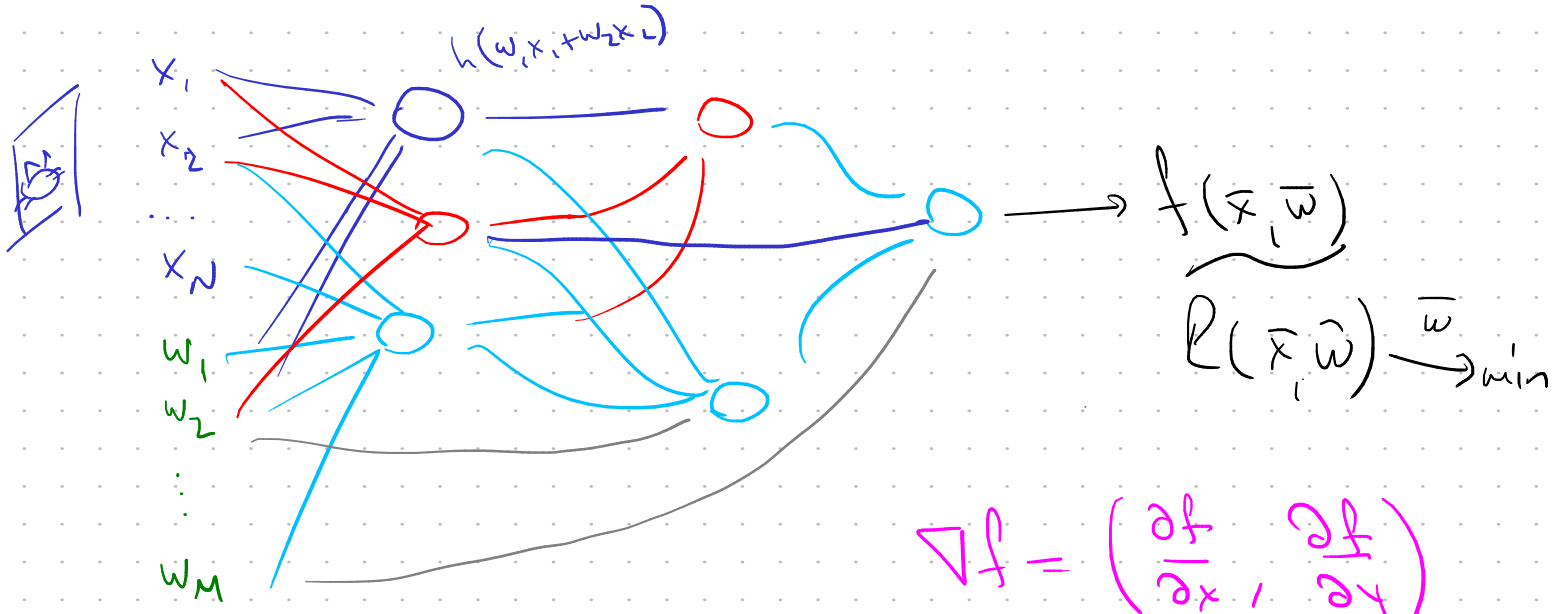
$p(C_k|\bar{x}) \in (0, \infty)$

$$\text{softmax}(a_1, a_2, \ldots, a_k) = \left(\ldots, \frac{e^{a_k}}{\sum_s e^{a_s}}, \ldots\right)$$



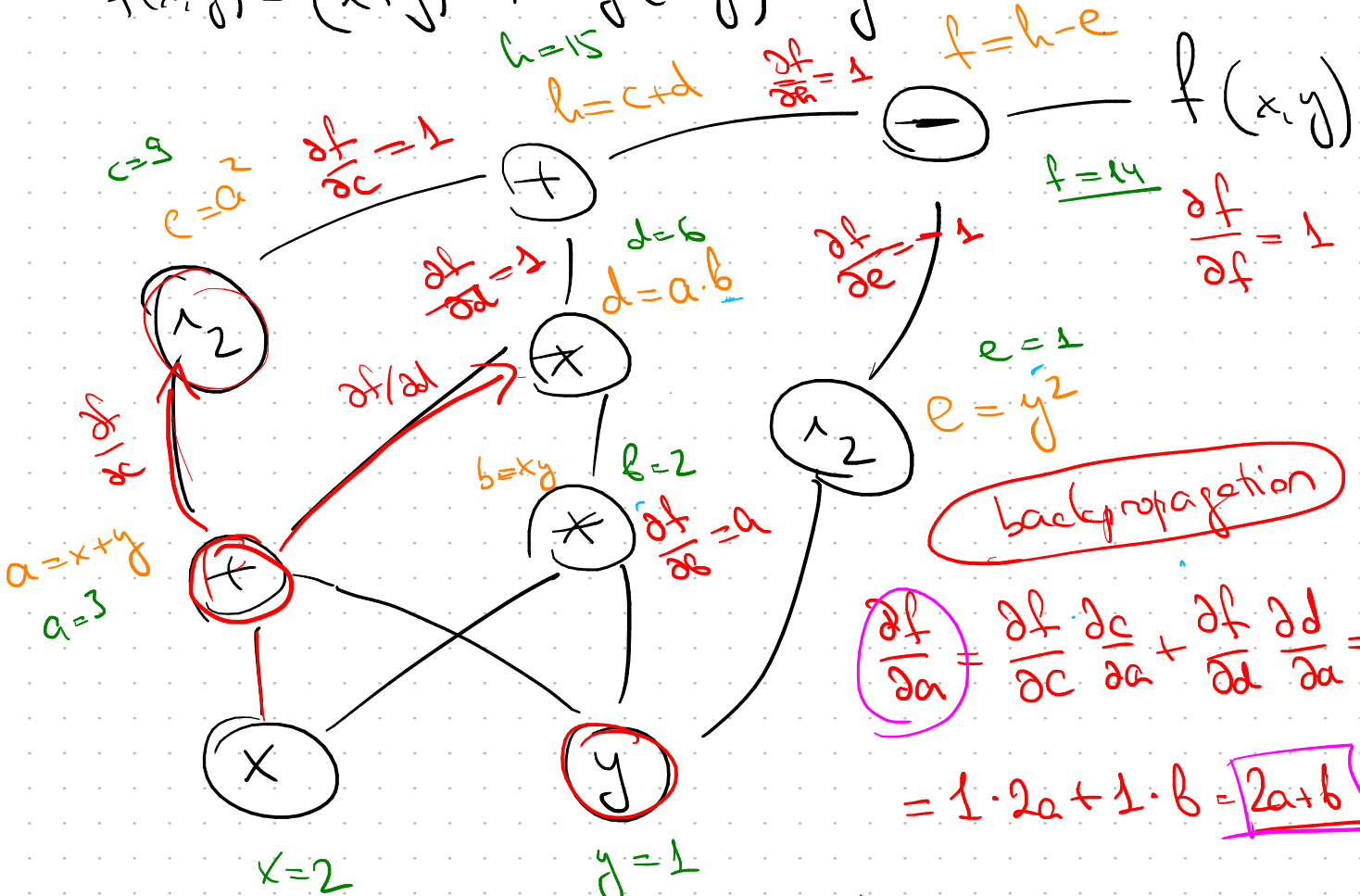


$$\text{ReLU}(x) = \max(0, x)$$



$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right)$$

$$f(x,y) = (x+y)^2 + xy(x+y) - y^2$$

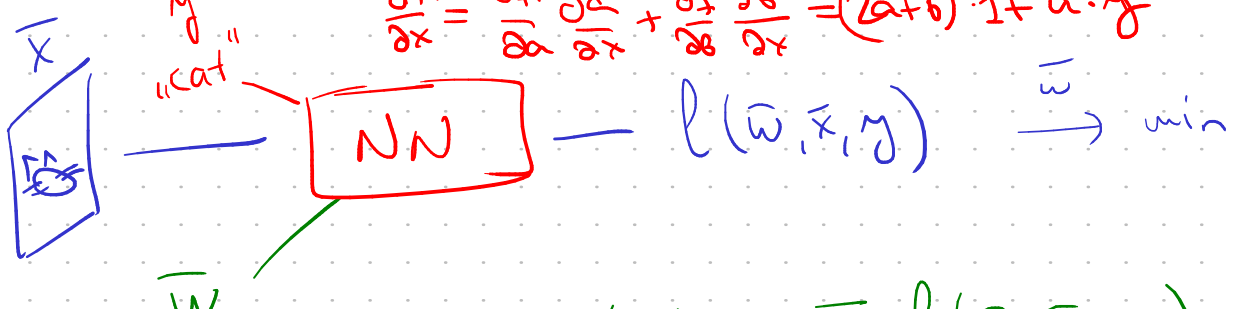


Backpropagation

$$\frac{\partial f}{\partial a} = \frac{\partial f}{\partial c}\frac{\partial c}{\partial a} + \frac{\partial f}{\partial d}\frac{\partial d}{\partial a} =$$

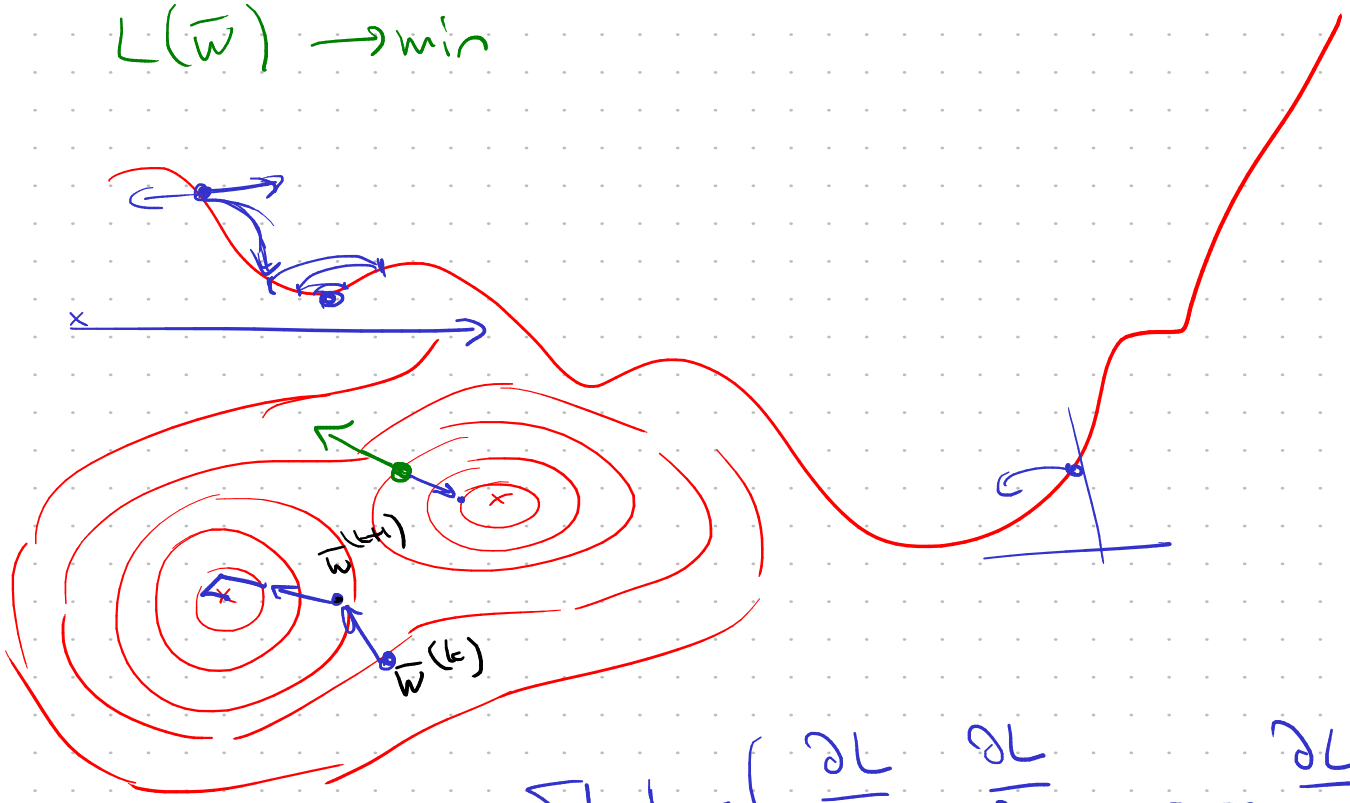
$$= 1 \cdot 2a + 1 \cdot b = 2a + b$$

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial a}\frac{\partial a}{\partial x} + \frac{\partial f}{\partial b}\frac{\partial b}{\partial x} = (2a+b)\cdot 1 + a \cdot y$$



$$L(\bar{w}) = \sum_n \ell(\bar{w}, \bar{x}_n, y_n)$$

$L(\bar{w}) \to \min$



$$\nabla_{\bar{w}} L = \left(\frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial w_2}, \dots, \frac{\partial L}{\partial w_n} \right)$$

$$\bar{w}^{(k+1)} := \bar{w}^{(k)} - \eta \cdot \nabla_{\bar{w}} L \Big|_{\bar{w}^{(k)}}$$

$$\frac{\partial f}{\partial y} = \frac{\partial f}{\partial a} \frac{\partial a}{\partial y} + \frac{\partial f}{\partial b} \cdot \frac{\partial b}{\partial y} + \frac{\partial f}{\partial e} \cdot \frac{\partial e}{\partial y} = 2a + b + a - x - 2y$$