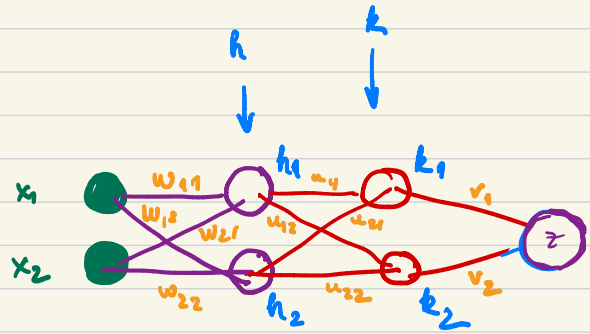
Thursday 02/22/2024



No bias

ReLU

$$h_1 = \mathrm{ReLU}(w_{11}x_1 + w_{21}x_2)$$
$$h_2 = \mathrm{ReLU}(w_{12}x_1 + w_{22}x_2)$$

1. $$h = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \mathrm{ReLU}\left(\underbrace{\begin{pmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \end{pmatrix}}_{W^T}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right)$$

$$W = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix}$$

- log loss
- binary - cross entropy
- logistic loss

$\mathrm{Loss}(y, p)$
$= -[y\log p + (1-y)\log(1-p)]$

2. $$k = \begin{pmatrix} k_1 \\ k_2 \end{pmatrix} = \boxed{\mathrm{ReLU}[U^T h]}$$

$$U = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix}$$

$$\boxed{\theta = (W, U, V)}$$

3. $V^T k$ $\qquad V = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$

(i.e logist) sigmoid

$$z = f(\theta, x) = V^T k \qquad\qquad z = f(\theta, x) = \sigma[V^T k]$$

<u>Regression problem</u>

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

want

$$f(\theta, x^{(i)}) \approx y^{(i)} \qquad \forall i = 1, \ldots, n$$

$$L(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left[f(\theta, x^{(i)}) - y^{(i)}\right]^2$$

<u>Classification (binary)</u>

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

$$y^{(i)} \in \{0, 1\}$$

$$L(\theta) = \frac{1}{n}\sum_{i=1}^{n} -[y^{(i)}\log(f(\theta, x^{(i)})) + (1 - y^{(i)})\log(1 - f(\theta, x^{(i)}))$$

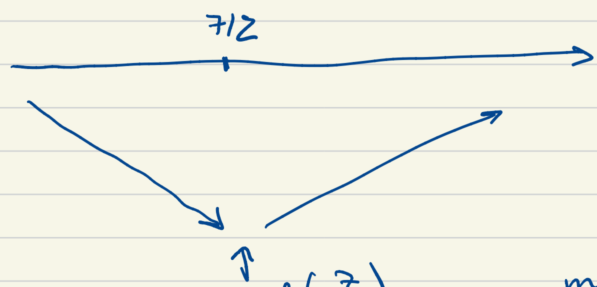
Calculus 1

① $$f(x) = x^2 - 7x + 9 \qquad x \in \mathbb{R}$$

$$f'(x) = 2x - 7$$

Critical points: $2x - 7 = 0$

$$x = \frac{7}{2}$$



$$f\left(\tfrac{7}{2}\right) = \min_x f(x)$$

① Start with $x = 0$

$f'(0) = -7 \rightarrow$ Increase x

② $\longrightarrow$ $x = 5$

$f'(5) = 3$

③ $\rightarrow$ $x = 4$

② Calculus 2:

$$f(x, y) = x^2 - 2xy + 2y^2 - 4y$$

$$\min_{x,y} f(x, y)$$

Gradient:

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

$$= (2x - 2y, \quad 2y - 4)$$

Critical points

$$\begin{cases} 2x - 2y = 0 \\ 2y - 4 = 0 \end{cases} \rightarrow (x, y) = (2, 2)$$