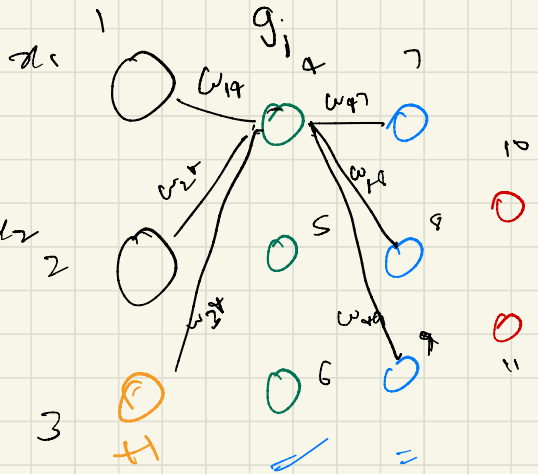
Input layer

Output layer




| | |
|---|---|
| $x_1$ | $y_1$, $l_1$ |
| $x_2$ | $y_2$, $l_2$ |
| $x_3$ | $y_3$, $l_3$ |
| ⋮ | |
| $x_n$ | $y_n$, $l_n$ |

Pytorch

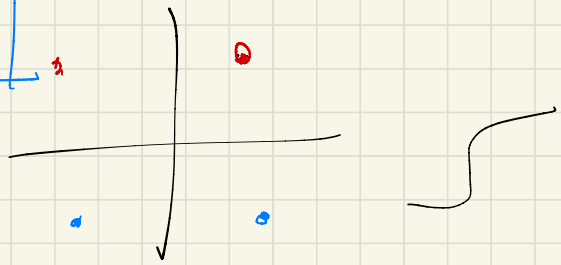
(1) Neural nets

(2) K-nn

$$in_4 = w_{14}x_1 + w_{24}x_2 + w_{34} \times (+1)$$

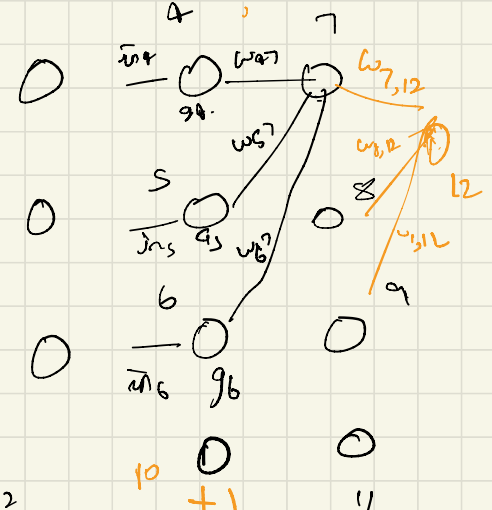
$$g_4(in_4) = \text{sigmoid}(w_{14}x_1 + w_{24}x_2 + w_{34})$$

$\Rightarrow$ activation function (non-linear fns)



- $\text{Sigmoid}(x) = \frac{1}{1+e^{-x}}$
- $\tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$
- $\text{ReLU} = \max(0, x)$

$$in_7 = \omega_{47}\, g_4(in_4) + \omega_{57}\, g_5(in_5) + \omega_{67}\, g_6(in_6) + \omega_{(0,7)}$$




$g_7(in_7)$

$$out = \omega_{7,12}\, g_7(in_7) + \omega_{8,12}\, g_8(in_8) + \omega_{9,12}\, g_9(in_9) + \omega_{11,12}$$

$\hookrightarrow \hat{y}$ $\qquad$ $y$

$\max(0, x)$

$$Error = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \qquad y$$

$\omega \to L$

$$\omega \leftarrow \omega - \eta \cdot \frac{\partial L}{\partial \omega}$$

$$\omega_{14} \leftarrow \omega_{14} - \eta \cdot \frac{\partial L}{\partial \omega_{14}}$$

Gradient descent.

$w_1$
$w_2$
$\vdots$
$w_n$

$$w \leftarrow w - \eta \nabla_w f$$

---

$$L = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

$$\nabla L = \nabla \left( \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \right)$$

$$= \sum_{i=1}^{n} \nabla (\hat{y}_i - y_i)^2$$

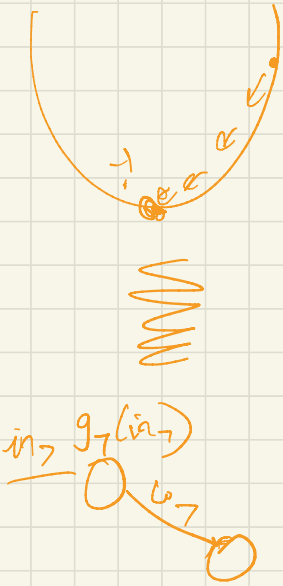
$\nabla_w (\hat{y} - y)^2$

$\hat{y}(w_1)$

$\frac{\partial L}{\partial w_1} = \frac{\partial}{\partial w_1} (\hat{y} - y)^2 =$

$2(\hat{y} - y) \cdot \frac{\partial}{\partial w_1} (\hat{y}) =$

$= 2(\hat{y} - y) \frac{\partial}{\partial w_1} \hat{y} = 2(\hat{y} - y) \cdot \frac{\partial}{\partial w_1} (w_1 g_1(w_1) + w_0 \cdot 1)$




$f(w)$

$\dot{w}$

$\nabla f \rightarrow$

$f(w + \nabla_w f)$

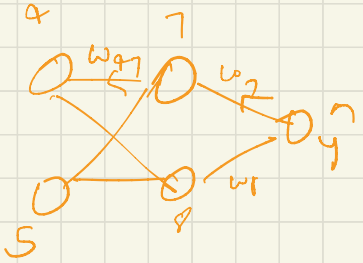
$f(w + t \cdot v)$

As $t \to 0$

$f(w+tv) - f(w)$

$= ((\nabla_w f)^T v) t$

$$\frac{\partial L}{\partial w_7} = 2(\hat{y}-y)\, g_7(in_7)$$




$$\frac{\partial L}{\partial w_{47}} = \frac{\partial}{\partial w_{47}}\left((\hat{y}-y)^2\right)$$

$$= 2(\hat{y}-y)\,\frac{\partial}{\partial w_{47}}\left(w_7\, g_7(in_7) + w_8\, g_8(in_8)\right)$$

$$= 2(\hat{y}-y) \cdot w_7 \cdot g_7'(in_7) \times \frac{\partial}{\partial w_{47}}\left(w_{47}\, g_4(in_4) + w_{57}\right)$$

$$= 2(\hat{y}-y) \times w_7\, g_7'(in_7) \times g_4(in_4)$$

$g \Rightarrow \frac{1}{1+e^{-x}}$

$g' = \frac{-1}{(1+e^{-x})^2}(-e^{-x})$

$= \frac{e^{-x}}{(1+e^{-x})^2}$

- Chapter 22 of Russell & Norvig AI book

22.1

- pytorch