⇒ Deep learning is a ... 6

→

Page 3 - review paper

Deep learning
- Traditional machine learning
- Age of Big data

Perceptron
→ forward propogation
→ $w_i$ --- $w_m$ (weight)
→ bias shift to left to right

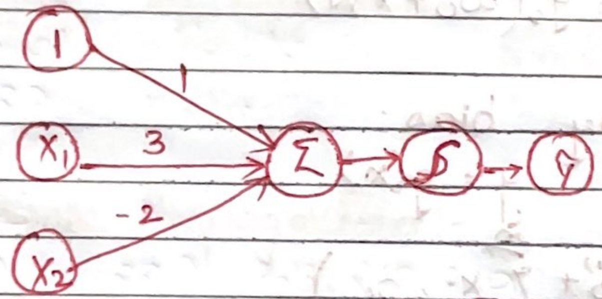
$$Y = g\left(w_0 + \sum_{i=1}^{m} x_i w_i\right)$$

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ \vdots \\ w_m \end{bmatrix}$$

⇒ non-linear activation function
→ $\sigma(x) = \frac{1}{1+e^{-x}}$

Activation function
→ To introduce non-linear activities

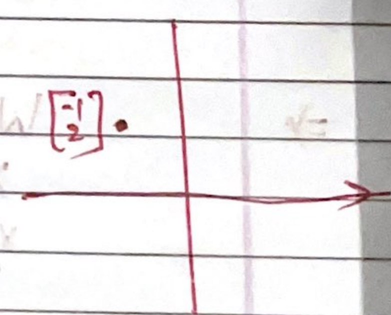
$\hat{y} = g(w_0 + X^T W)$




$w_0 = 1 \qquad W = \begin{bmatrix}3\\-2\end{bmatrix}$

$\hat{y} = g\left[1 + \begin{bmatrix}x_1\\x_2\end{bmatrix}\begin{bmatrix}3\\-2\end{bmatrix}\right]$

$= g\left[1 + 3x_1 - 2x_2\right]$




If $(-1, 2) \Rightarrow \quad X = \begin{bmatrix}-1\\2\end{bmatrix}$

$\Rightarrow g(-6) = 0.002$ ; Sigmoid function

=0>1

>0 = 0.5

⇒ Now Scaling up size of n/w <0 = 0.5

Building Neural n/w

$$\hat{y} = g_0(w_0 + x^T W)$$

Take away bias.

$$y = g(x)$$

$$Z = w_0^{(1)} + \sum_{j=1}^{m} x_j w_{j,1})$$

$$(w_{0,i}^{(2)}$$

- Will I pass the Class!

$x_1$ = Number of lectures you attend

$x_2$ = Hours spent on final project

$$= \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

$$x' = [4, 5]$$

$$\underbrace{L(f(x^{i}; W)}_{\text{Predicted}}, \underbrace{y^{(i)}}_{\text{Actual}}) \quad L = \text{Loss}$$

$$J(W) = \frac{1}{n} \sum_{i=1}^{n} L(f(x^{(i)}; W), y^{(i)})$$

Mean Squared Error Loss.

$$W = \arg\min \frac{1}{n} \sum_{i=1}^{n} L\left(f(x^{(i)}; W), y^{(i)}\right)$$

Loss optimization :-

gradient descent

Back Propogation :-

$$x \xrightarrow{w} x_1 \xrightarrow{w_2} \hat{y} \longrightarrow J(w)$$

$$\Rightarrow \frac{dJ(w)}{w_2} = \frac{df(w)}{d\hat{y}} \times \frac{d\hat{y}}{dw_2}$$

$$\frac{dJ(w)}{w_1} = \frac{df(w)}{dZ} \times \frac{dZ_1}{dw_1}$$

$$\Rightarrow W - \eta \frac{dJ(w)}{dw}$$

Learning rate setting is challenging.