16 July Day 2

22 November 2022 06:57

$\Sigma = w_1 x_1 + w_2 x_2 + b$

$\text{sig}(z) = \frac{1}{1+e^{-z}}$

Output = $\boxed{0 \text{ to } 1}$

① student pass 1 (OR GATE)

play	carry	class
✓ 0	0	0
✓ 0	1	1
1	0	1
1	1	1

threshold (boundary)

If $\boxed{1p} = \text{class } 1$

$\boxed{1c} = \text{class } 0$

② student (AND gate)

play	carry	class
0	0	0
0	1	0
1	0	0
1	1	1

threshold

≥ 2 class 1

<2 class 0

③ student (XOR gate)

play	carry	class
0	0	0
0	1	1
1	0	1
1	1	0

threshold

= class 0

= class 1

③ not gate represent non linearly separable data.

④ OR, AND gate represent linearly separable data.

that was why we can write classical perceptron was able to work on linearly separable data but not on non linearly separable data.

$\boxed{\text{MLN}}$

x_1	x_2
0	0
0	1
1	0
1	1

no of N_1 $\boxed{0}$ complexity

no of N_2 $\boxed{0}$ pattern together not able to ↗

$y = mx + c$

$y = 0x + c$

output

x is constant

variable parameter

= slope, intercept

weights

It is justifying importance of feature with respect to target variable.

Bias

It is an adjusting factor.

$y = 0x + c$ $\quad 100 = \overset{0}{m}x + c$

$100 = 0 + c$

$100 = 0 + c$

$100 = 0 + c$

$100 = 0 + c$

$100 = 0 + c$

$100 \quad c$

Your $\rightarrow 100 = 80 + \hat{y}$

$\text{your} - \hat{y} \Rightarrow \boxed{\text{loss}} = \boxed{20}$

$\text{your} - \hat{y} = 100 - 80$

$\boxed{\text{loss} = 0}$

$\text{new} = \text{old} - \alpha \frac{\partial L}{\partial \text{old}}$ slope

$\text{new} = \text{old} - \alpha \frac{\partial L}{\partial \text{old}}$ intercept

Chain Rule

forward propagation

back propagation

$\text{new}_{\text{old}} = \text{old} - \alpha \frac{\partial L}{\partial \text{old}}$

$w_{11}^{2}\text{new} = w_{11}^{2}\text{old} - \alpha \frac{\partial L}{\partial w_{11}^{2}\text{old}}$

$\frac{\partial L}{\partial w_{11}^{2}\text{old}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_{31}} \cdot \frac{\partial z_{31}}{\partial w_{11}^{2}\text{old}}$

$w_{11}^{2}\text{new} = w_{11}^{2}\text{old} - \alpha \frac{\partial L}{\partial w_{11}^{2}\text{old}}$

input layer $\quad$ HL 1 $\quad$ HL 2 $\quad$ output layer

$\boxed{4 \times 3}$ $\quad$ $\boxed{3 \times 2}$ $\quad$ $\boxed{2 \times 1}$

$\frac{\partial L}{\partial w_{21}^{2}\text{old}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_{31}} \cdot \frac{\partial z_{31}}{\partial w_{21}^{2}\text{old}}$

weights 1

$$\begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \\ w_{41} & w_{42} & w_{43} \end{bmatrix}$$

$w_{11}^{1}\text{new} = w_{11}^{1}\text{old} - \alpha \frac{\partial L}{\partial w_{11}^{1}\text{old}}$

$\frac{\partial L}{\partial w_{11}^{1}\text{old}} = \left[\frac{\partial L}{\partial w_{11}^{2}} \cdot \frac{\partial w_{11}^{2}}{\partial z_{11}} \cdot \frac{\partial z_{11}}{\partial w_{11}^{1}\text{old}} \right]$

weights 2

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \end{bmatrix}$$

$+ \left[\frac{\partial L}{\partial w_{11}^{2}} \cdot \frac{\partial w_{11}^{2}}{\partial z_{11}} \cdot \frac{\partial z_{11}}{\partial w_{11}^{2}\text{old}} \right]$

weights 3

$$\begin{bmatrix} w_{11} \\ w_{21} \end{bmatrix}$$

Activation fn are very important

Gradient Descent

$y = mx + c$

trainable parameter $\boxed{m, c}$

slope is having relational impact

intercept is having vertical impact

(A) −ve slope

(B) +ve slope.

weights

$\text{Loss} = \sum_{i=1}^{n} (y - \hat{y})^2$

$= (y_{\text{act}} - \hat{y})^2$

$\text{Loss} = (y_{\text{act}} - (mx + c))^2$

learning rate.

step size

convergence point

Global minima.

weight

Loss

slope 0

weight → $\hat{y}$ → Loss → $\boxed{\text{BP}}$ → Repeat

w, B (A)

$\boxed{\alpha = 0.01}$

gradient descent

gradient random descent

Loss $\boxed{1}$ $\quad$ epoch $\boxed{1}$

when to stop?

① early stopping $\boxed{\text{epoch}} = 10$

② when two gradient are equal to old gradient it stops for algorith.

$\boxed{\text{new} = \text{old}}$

10 ← 10

$\text{new} = \text{old} - \alpha \frac{\partial L}{\partial \text{old}}$

$\boxed{\text{new} = \text{old}}$

① pass the dataset

$\hat{y}$ $\quad$ Loss $\quad$ your − $\hat{y}$

w, B update

② pass the dataset on updated weights and biases

$\hat{y}$ $\quad$ Loss your − $\hat{y}$

w, B update