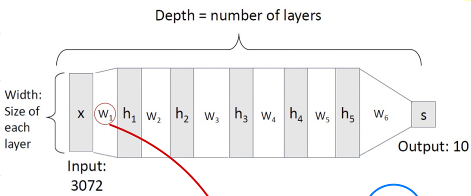
20222385 김우혁

Back propagation

합성함수의 미분 ex) $f \cdot g(x) \Rightarrow g'(x) \cdot f'(g(x))$

$= \frac{df}{\cancel{dg}} \cdot \frac{\cancel{dg}}{dx} = \frac{df}{dx}$



Q. 이러한 DNN에서 $w_1 = w_1 - \eta \nabla L$

바로 구하기 어려움

∴ Chain rule 이용

Gradient Descent

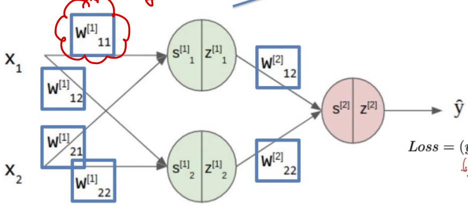
Method: at each step, move in direction of negative gradient

```
w0 = initialize()                    #initialize
for iter in range(numIters):
    g = ∇_w L(w)                     # eval gradient
    w = w + -stepsize*g              # update w
return w
```

Learning rate

예시로 이 녀석에 대한 gradient 찾아보자!

Learnable Parameters



$Loss = (y - \hat{y})^2$

gt

Then, How to calculate gradient on each layer ?

$$\frac{\partial Loss}{\partial w_{11}^{(1)}}, \frac{\partial Loss}{\partial w_{12}^{(1)}}, \dots, \frac{\partial Loss}{\partial w_{22}^{(2)}},$$

우리의 목표.

$$s_1^{[1]} = w_{11}^{[1]} x_1 + w_{21}^{[1]} x_2$$

$$z_1^{[1]} = tanh(s_1^{[1]})$$

$$s^{[2]} = w_{12}^{[2]} z_1^{[1]} + w_{22}^{[2]} z_2^{[1]}$$

$$z^{[2]} = tanh(s^{[2]})$$

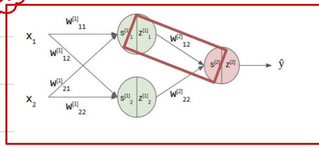
$$L = (y - z^{[2]})^2$$

gt (정답)

역으로 계산해 올라가야 함

$$\frac{\partial L}{\partial w_{11}^{[1]}} = \frac{\partial L}{\partial z^{[2]}} \times \frac{\partial z^{[2]}}{\partial s^{[2]}} \times \frac{\partial s^{[2]}}{\partial z_1^{[1]}} \times \frac{\partial z_1^{[1]}}{\partial s_1^{[1]}} \times \frac{\partial s_1^{[1]}}{\partial w_{11}^{[1]}}$$

주의



$$s^{[2]} = w_{12}^{[2]} z_1^{[1]} + w_{22}^{[2]} z_2^{[1]}$$

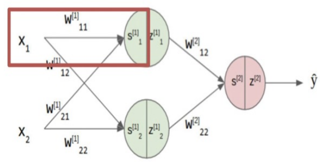
$$\frac{\partial s^{[2]}}{\partial z_1^{[1]}} = w_{12}^{[2]}$$

$$\frac{\partial L}{\partial w_{11}^{[1]}} = \frac{\partial L}{\partial z^{[2]}} \times \frac{\partial z^{[2]}}{\partial s^{[2]}} \times \frac{\partial s^{[2]}}{\partial z_1^{[1]}} \times ..$$

여기서, $z_1^{[1]}$으로 미분해줘야 chain rule 이용가능

∵ $z_1^{[1]} = tanh(s_1^{[1]})$

$w_{11}^{[1]}$에 도달할 때까지 이어서 전개 가능



$$s_1^{[1]} = w_{11}^{[1]} x_1 + w_{21}^{[1]} x_2$$

$$\frac{\partial s_1^{[1]}}{\partial w_{11}^{[1]}} = x_1$$

↳ 여기는 $w_{11}^{[1]}$으로 미분.

★ $$\frac{\partial L}{\partial w_{11}^{[1]}} = \boxed{\frac{\partial L}{\partial z^{[2]}}} \times \boxed{\frac{\partial z^{[2]}}{\partial s^{[2]}}} \times \boxed{\frac{\partial s^{[2]}}{\partial z_1^{[1]}}} \times \boxed{\frac{\partial z_1^{[1]}}{\partial s_1^{[1]}}} \times \boxed{\frac{\partial s_1^{[1]}}{\partial w_{11}^{[1]}}}$$

각각의 값을 역방향으로 전개하면서 구할 수 있음