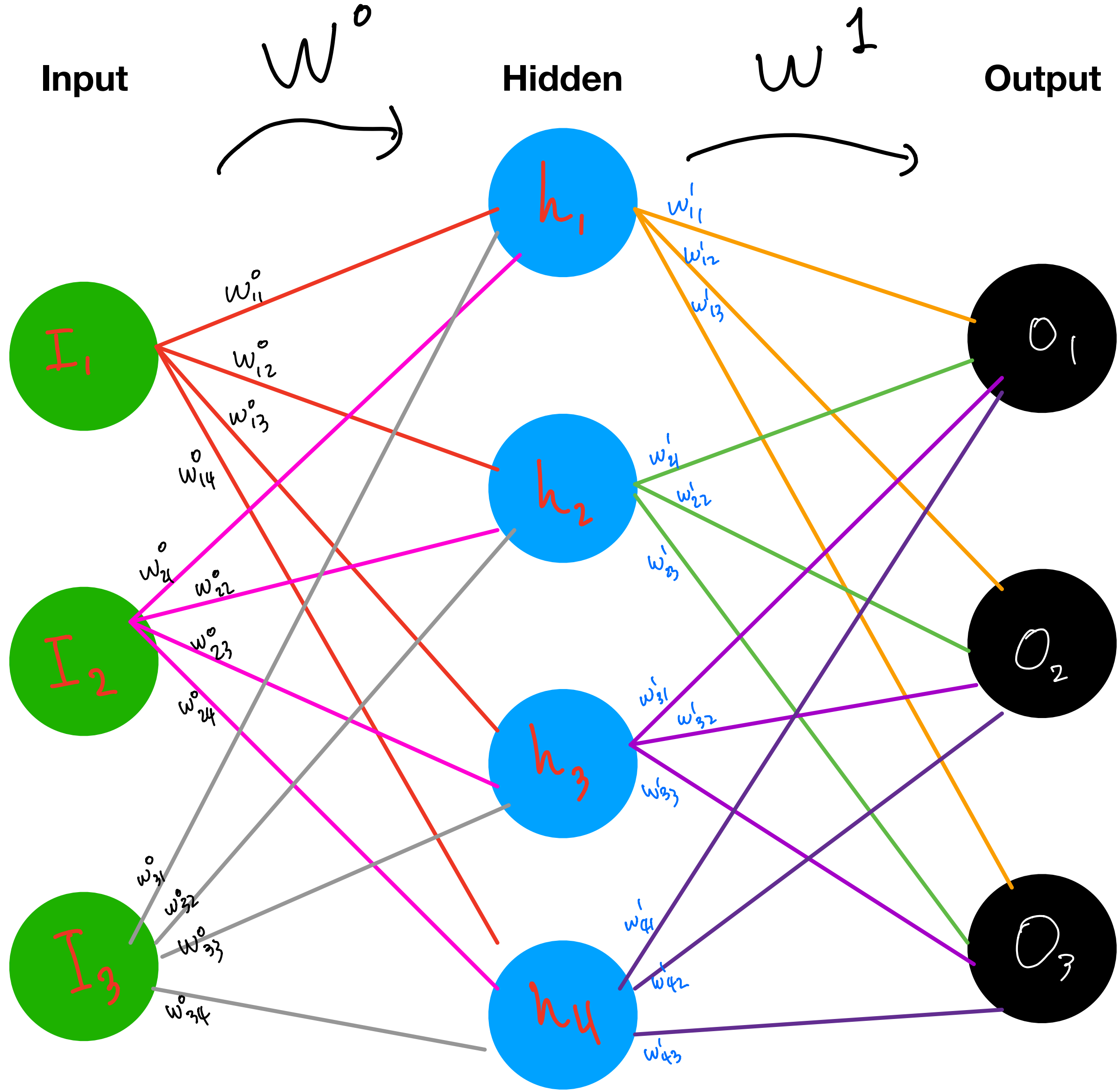


Input
W^0
Hidden
W^1
Output
I_1
I_2
I_3
h_1
h_2
h_3
h_4
O_1
O_2
O_3
w^0_{11}
w^0_{12}
w^0_{13}
w^0_{14}
w^0_{21}
w^0_{22}
w^0_{23}
w^0_{24}
w^0_{31}
w^0_{32}
w^0_{33}
w^0_{34}
w^1_{11}
w^1_{12}
w^1_{13}
w^1_{21}
w^1_{22}
w^1_{23}
w^1_{31}
w^1_{32}
w^1_{33}
w^1_{41}
w^1_{42}
w^1_{43}

기본구조 : feed.

$$w^0 = \begin{bmatrix} w^0_{11} & w^0_{12} & w^0_{13} & w^0_{14} \\ w^0_{21} & w^0_{22} & w^0_{23} & w^0_{24} \\ w^0_{31} & w^0_{32} & w^0_{33} & w^0_{34} \end{bmatrix}$$

#

3 X 4 matrix

$\underbrace{3}_{\text{Input}}$ X $\underbrace{4}_{\#\text{hidden}}$

$$h_1 = w^0_{11} I_1 + w^0_{21} I_2 + w^0_{31} I_3$$

$$h_2 = w^0_{12} I_1 + w^0_{22} I_2 + w^0_{32} I_3$$

$$h_3 = w^0_{13} I_1 + w^0_{23} I_2 + w^0_{33} I_3$$

$$h_4 = w^0_{14} I_1 + w^0_{24} I_2 + w^0_{34} I_3$$

$$h_j = \sum_{i=1}^{N} w^0_{ij} I_i$$

$N = \#$hidden

$$w^1 = \begin{bmatrix} w^1_{11} & w^1_{12} & w^1_{13} \\ w^1_{21} & w^1_{22} & w^1_{23} \\ w^1_{31} & w^1_{32} & w^1_{33} \\ w^1_{41} & w^1_{42} & w^1_{43} \end{bmatrix}$$

4 x 3 matrix

$\underbrace{4}_{\#\text{hidden}}$ x $\underbrace{3}_{\#\text{output}}$

$$O_1 = w^1_{11} h_1 + w^1_{21} h_2 + w^1_{31} h_3 + w^1_{41} h_4$$

$$O_2 = w^1_{12} h_1 + w^1_{22} h_2 + w^1_{32} h_3 + w^1_{42} h_4$$

$$O_3 = w^1_{13} h_1 + w^1_{23} h_2 + w^1_{33} h_3 + w^1_{43} h_4$$

$$O_j = \sum_{i=1}^{M} w^1_{ij} h_i$$

$M = \#$output

$$h = np.dot[I, W^0]$$

$$\boxed{z_1 = h}$$

$$O = np.dot[h, w^1]$$

여기서

넘어갈때

Activation(활성화 함수를 추가해줌)

∵ 결과에 Nonlinearity (비선형성) 성질을 부여하기위해.

우리는 ReLU 함수 쓸 거야 ReLU(h)

그래서 실은 $O = np.dot[Act[h], w^1]$

$$\boxed{z_2 = O}$$

//

Cost : output(prediction) vs true label..

얼마나 맞았는지 비교해주세요.

$\Rightarrow -\Sigma(\text{true} * \log(\text{prediction}))$: cross entropy

formally: $Cost = -\sum_j y_j \log P_j$, y_j: true label
P_j: prediction

∨ 우리의 문제 해결방식은 output 값이 어떤 class로 분류되는지 궁금하다.

그래서 0~9로 나누어진 10개의 클래스중 어느곳에 속할 "확률"이 가장 높은지 알아내야 하기 때문에 output을 합이 1인 확률벡터로 변환 하기위해 Softmax 함수를 사용한다.

Softmax : $\sigma(z)_i = \dfrac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$ for $i=1,\cdots,K$ and $z=(z_1,\cdots,z_K) \in \mathbb{R}^K$

우리의 output.

최종적으로 $\sigma(z)_i$ 가 우리의 prediction 임.

그래서 일단 random한 $W^{(0)}, W^{(1)}$로 prediction 까지 다 해주면.
이제 정확도를 올리기위한 작업을 해줘야 하는데..

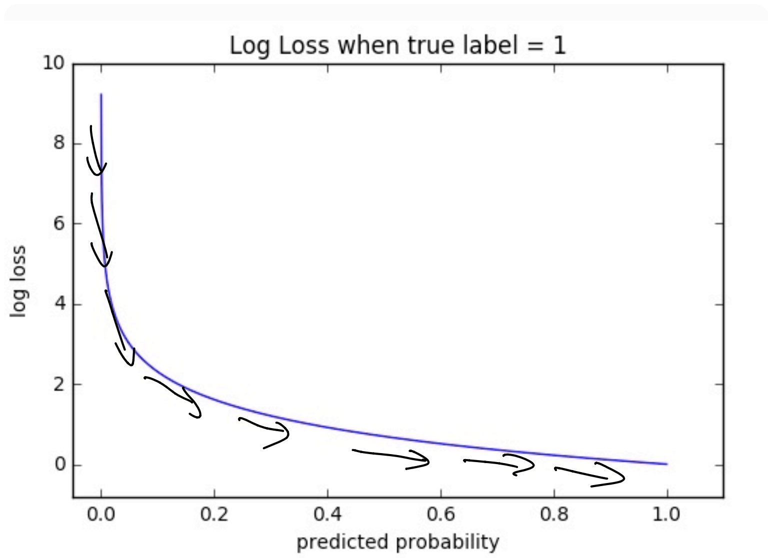
이를 하여

Back Propagation

$w^{(0)}$, $w^{(1)}$을 특정한 방식을 통해 새롭게 update를 해줘야 만해!

어떻게 할건데? Cost 함수(다른정도)를 0으로 만들어주자

Cross entropy



Cost 함수의 변화량을 $w^{(0)}, w^{(1)}$ update 할때 반영 해보자!

gradient

Cost를 미분해볼까 뭐에 대한 변화량을 알고싶은데? $\boxed{w^{(0)}, w^{(1)}}$ 이 변화함에따라 변하는 Cost를 보고 싶어!

우선 거꾸로 바꿔가야 하니 $w^{(1)}$에 대해서 생각해보자!

$$\begin{aligned}
Cost &= -\sum_j y_j \log Prediction_j \\
&= -\sum_j y_j \log(Softmax(output)) \\
&= -\sum_j y_j \log(softmax(\overbrace{Relu(hidden) \times w^{(1)}}^{z_2}))
\end{aligned}$$

$$\frac{\partial Cost}{\partial w^{(1)}} = \frac{\partial cost}{\partial z_2}\frac{\partial z_2}{\partial w^{(1)}} = \delta^2 \times Relu(hidden)$$

$$\delta^2 = \frac{\partial \, Cost}{\partial z_2} = \frac{\partial \; - \sum_j y_j \log(Softmax(z_2))}{\partial \; z_2}$$

let

$$p(z_2)_i = softmax(z_2)_i = \frac{e^{(z_2)_i}}{\sum_{j=1}^{K} e^{(z_2)_j}}$$

$$\frac{\partial \, cost}{\partial z_2} = - \sum_j y_j \frac{\partial \log p(z_2)_j}{\partial z_2}$$

$$= - \sum_j y_j \frac{1}{p(z_2)_j} \frac{\partial p(z_2)_j}{\partial z_2}$$

$$= - \left(y_i(1 - p(z_2)_i) + \sum_{j \neq i} y_j \frac{1}{\cancel{p(z_2)_j}} \left(- \cancel{p(z_2)_j} \, p(z_2)_i \right) \right)$$

$$= - \left(y_i(1 - p(z_2)_i) - \sum_{j \neq i} y_j (p(z_2)_i) \right)$$

$$= - \left(y_i - y_i p(z_2)_i - \sum_{j \neq i} y_j (p(z_2)_i) \right)$$

$$= - \left(y_i - \sum_j y_j p(z_2)_i \right)$$

$$= \left(\underline{\sum_j y_j p(z_2)_i} - y_i \right)$$

$$= p(z_2)_i \Big(\underbrace{\sum_j y_j}_{1} \Big) - y_i = \underline{\underline{softmax(z_2) - true}}$$

$$\therefore \frac{\partial Cost}{\partial W^{(1)}} = (\text{Softmax}(z_2) - \text{true}) \times \underline{\text{ReLU}(\text{hidden})}.$$

이제 $w^{(0)}$ 에 대한 변화량도 구해보자.

$$Cost = - \sum_j y_j \log \text{Prediction}_j$$

$$= - \sum_j y_j \log(\text{Softmax}(\text{output}))$$

$$= - \sum_j y_j \log(\text{softmax}(\underbrace{\text{Relu}(\text{hidden}) \times w^{(1)}}_{z_2}))$$

$$= - \sum_j y_j \log(\text{softmax}(\text{ReLU}(\underbrace{w^{(0)} x + b}_{z_1}) \times w^{(1)}))$$

$$\frac{\partial Cost}{\partial w^{(0)}} = \frac{\partial cost}{\partial z_1} \cdot \frac{\partial z_1}{\partial w^{(0)}} = \delta^{(1)} \cdot \underline{x} \rightarrow \text{Input}.$$

$$\frac{\partial cost}{\partial z_1} = \frac{\partial cost}{\partial \text{ReLU}(z_1)} \cdot \frac{\partial \text{ReLU}(z_1)}{\partial z_1}$$

$$= \frac{\partial cost}{\partial z_2} \cdot \frac{\partial z_2}{\partial \text{ReLU}(z_1)} \cdot \frac{\partial \text{ReLU}(z_1)}{\partial z_1}$$

$$= \underline{\delta^{(2)}} \cdot \underline{w^{(1)}} \cdot \underline{(z_1 > 0)}$$

ReLU 미분하면 양수인부분만 1 나머지는 0

$$\therefore \frac{\partial Cost}{\partial w^{(0)}} = \underbrace{\delta^{(2)} \cdot w^{(1)} \cdot (z_1 > 0)}_{\delta^{(1)}} \cdot x$$