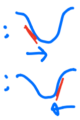
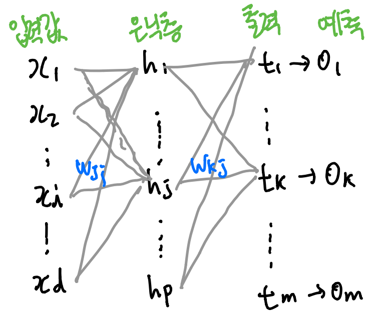
김성범 교수님 backpropagation 명강 ~!

[경사하강법] $W_{t+1} = W_t - \alpha \cdot \underbrace{L'(W_t)}$ $(0<\alpha<1$: learning rate$)$

<0 :

>0 :





입력 출력 예측

$D_1 = (x_{11}, x_{12}, x_{13}, \cdots x_{1d}, t_{11}, t_{12}, \cdots t_{1m})\ (o_{11}, o_{12}, \cdots, o_{1m})$

$D_2 = (x_{21}, x_{22}, \cdots x_{2d}, t_{21}, t_{22}, \cdots t_{2m})\ (o_{21}, o_{22}, \cdots o_{2m})$

$\vdots$

$D_N = (x_{N1}, x_{N2}, \cdots x_{Nd}, t_{N1}, t_{N2}, \cdots t_{Nm})\ (o_{N1}, o_{N2}, \cdots o_{Nm})$

관측치 N개

※ 아래 정의된 비용함수를 최소로 하는 W를 찾자! (Y: 연속형일 때 예시)

$$E(W) = \sum_{n=1}^{N} E_n(W) \qquad E_n(W) = \frac{1}{2}\sum_{k=1}^{m}(t_{nk} - o_{nk})^2$$

※ $\frac{\partial E_n}{\partial w_{kj}}$: 출력층과 은닉층 사이 / $\frac{\partial E_n}{\partial w_{ji}}$: 은닉층과 입력층 사이

$$h_j = \frac{1}{1 + \exp(-(w_{j0} + \sum_{i=1}^{d} w_{ji} \cdot x_i))}$$

$$o_k = \frac{1}{1 + \exp(-(w_{k0} + \sum_{j=1}^{p} w_{kj} \cdot h_j))}$$

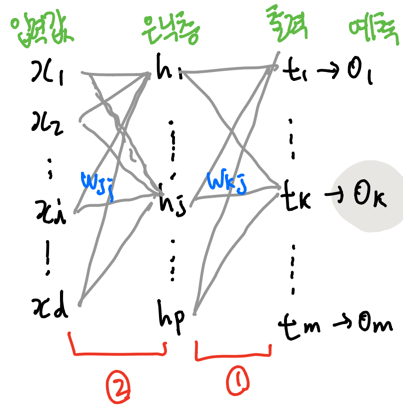
← 히든 노드의 선형결합을 시그모이드에 집어넣은거

$$\frac{1}{2}\sum_{n=1}^{m}\left(t_k - \frac{1}{1+\exp\left(-\left(w_{k0}+\sum_{j=1}^{p} w_{kj}\left(\frac{1}{1+\exp\left(-\left(w_{j0}+\sum_{i=1}^{d} w_{ji}\cdot x_i\right)\right)}\right)\right)\right)}\right)^2$$

바로미분하기 너무복잡

$$E_n(W) = \frac{1}{2}\sum_{k=1}^{m}(t_{nk}-O_{nk})^2$$

so, ① 출력과은닉 사이 → ② 은닉과 입력사이 순으로 미분 진행!



① 출력과은닉 사이

$net_k = h_1 \cdot w_{k1} + h_2 \cdot w_{k2} + h_3 \cdot w_{k3} + \cdots + h_j \cdot w_{kj} + \cdots + h_p \cdot w_{kp} + w_{k0}$

↘ 이때 t_k가 아님!

h_j: 은닉층의 j번째 노드값

$$O_k = \text{sigmoid}(net_k) = \frac{1}{1+\exp(-net_k)}$$

$\underset{\text{활성함수}}{=} \tau(net_k)$

$$E_n(W) = \frac{1}{2}\sum_{k=1}^{m}(t_k-O_k)^2$$

→ 여기 t_k는 출력층이 아니라 실제 y값! (뭔가 잘못된건지 유튜브댓글 달아놨으니 답글 달리면 고치셈!)

$$\Delta w_{kj} = -\alpha \frac{\partial E_n}{\partial w_{kj}}$$

$$\frac{\partial E_n}{\partial w_{kj}} = \frac{\partial E_n}{\partial net_k}\cdot\frac{\partial net_k}{\partial w_{kj}} = \frac{\partial E_n}{\partial O_k}\cdot\frac{\partial O_k}{\partial net_k}\cdot\frac{\partial net_k}{\partial w_{kj}} = \underset{①}{\frac{\partial E_n}{\partial O_k}}\cdot\underset{②}{\frac{\partial O_k}{\partial net_k}}\, h_j = \underbrace{(t_k-O_k)\overset{\tau'(net_k)}{O_k(1-O_k)}}_{\delta} h_j$$

미분하면

$net_k = \cancel{h_1 \cdot w_{k1}} + \cancel{h_2 \cdot w_{k2}} + \cancel{h_3 \cdot w_{k3}} + \cdots + h_j \cdot w_{kj} + \cdots + \cancel{h_p \cdot w_{kp}} + \cancel{w_{k0}}$

w_{kj}의 계수인 h_j만 남고 나머지 전부 사라짐

$$① \ \frac{\partial E_n}{\partial O_k} = \frac{\partial}{\partial O_k}\frac{1}{2}\sum_{k=1}^{m}(t_k-O_k)^2 = \frac{\partial}{\partial O_k}\frac{1}{2}(t_k-O_k)^2 = -(t_k-O_k)$$

O_k에 관련된 부분만 보면 돼서 Σ 사라짐

겉미분이미

$$② \ \frac{\partial O_k}{\partial net_k} = \frac{\partial}{\partial net_k}\frac{1}{1+\exp(-net_k)} = \frac{\exp(-net_k)}{(1+\exp(-net_k))^2} = O_k(1-O_k)$$

$= \tau'(net_k)$

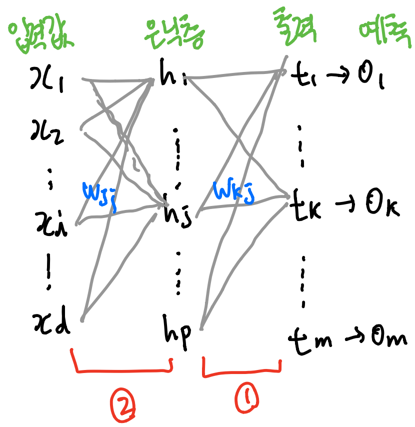
$$\therefore \Delta w_{kj} = -\alpha\frac{\partial E_n}{\partial w_{kj}} = \alpha(t_k-O_k)O_k(1-O_k)h_j$$

$$= \alpha(t_k-O_k)\tau'(net_k)h_j$$

→ 활성함수가 sigmoid가 아니라면 바뀔수있는부분임!

$$w_{kj} = w_{kj} - \alpha\frac{\partial E_n}{\partial w_{kj}}$$

② 은닉층과 입력층 사이



$$net_j = x_1 W_{j1} + x_2 W_{j2} + \cdots + x_i W_{ji} + \cdots + x_d W_{jd} + W_{j0}$$

$$h_j = \mathrm{sigmoid}(net_j) = \frac{1}{1+\exp(-net_j)}$$

$$= \tau(net_j)$$

$$net_k = h_1 \cdot W_{k1} + h_2 \cdot W_{k2} + \cdots + h_j \cdot W_{kj} + \cdots + h_p \cdot W_{kp} + W_{k0}$$

$$O_k = \mathrm{sigmoid}(net_k) = \frac{1}{1+\exp(-net_k)}$$

$$= \tau(net_k)$$

$$E_n(W) = \frac{1}{2}\sum_{k=1}^{m}(t_k - O_k)^2, \quad \Delta W_{ji} = -\alpha \frac{\partial E_n}{\partial W_{ji}}$$

$$\frac{\partial E_n}{\partial W_{ji}} = \frac{\partial E_n}{\partial net_j} \cdot \frac{\partial net_j}{\partial W_{ji}} = \frac{\partial E_n}{\partial net_j} \cdot x_i$$

미분하면

$$net_j = \cancel{x_1 W_{j1}} + \cancel{x_2 W_{j2}} + \cdots + x_i W_{ji} + \cdots + \cancel{x_d W_{jd}} + \cancel{W_{j0}}$$

W_{ji}의 계수인 x_i만 남고 다사라짐

$$\frac{\partial E_n}{\partial net_j} = \frac{\partial}{\partial net_j} \frac{1}{2}\sum_{k=1}^{m}(t_k - O_k)^2 = \frac{1}{2}\sum_{k=1}^{m}\frac{\partial}{\partial net_j}(t_k - O_k)^2$$

$$= \frac{1}{2}\sum_{k=1}^{m} \underbrace{\frac{\partial h_j}{\partial net_j}}_{①} \underbrace{\frac{\partial net_k}{\partial h_j}}_{②} \underbrace{\frac{\partial O_k}{\partial net_k}}_{③} \underbrace{\frac{\partial (t_k - O_k)^2}{\partial O_k}}_{④}$$

① $\dfrac{\partial h_j}{\partial net_j} = h_j(1-h_j) = \tau'(net_j)$ ② $\dfrac{\partial net_k}{\partial h_j} = W_{kj}$ ③ $\dfrac{\partial O_k}{\partial net_k} = O_k(1-O_k) = \tau'(net_k)$ ④ $\dfrac{\partial (t_k - O_k)^2}{\partial O_k} = -2(t_k - O_k)$

$$= -\underbrace{h_j(1-h_j)}_{\tau'(net_j)} \sum_{k=1}^{m} W_{kj} \underbrace{O_k(1-O_k)}_{\tau'(net_k)}(t_k - O_k)$$

왜 가중치가 곱해지는지 알겠지!?

$$\therefore \Delta W_{ji} = -\alpha \frac{\partial E_n}{\partial W_{ji}} = \alpha x_i \underbrace{h_j(1-h_j)}_{\tau'(net_j)} \boxed{\sum_{k=1}^{m} W_{kj} \underbrace{O_k(1-O_k)}_{\tau'(net_k)}(t_k - O_k)}$$

$$W_{ji} = W_{ji} - \alpha \frac{\partial E_n}{\partial W_{ji}}$$

실습코드에서는 이친구를 error로 보는거이고

errors = [$error_0$, $error_1$, ⋯ $error_j$, ⋯]

j번째 error와 2때의 $\tau'(net_j)$가 곱해져서 새로운 δ(델타) 생성