심화 학습

<오차 역전파의 이해>

1. 오차 역전파의 설계

오차 역전파 구동 방식은 다음과 같이 정리될 수 있습니다.

1) 임의의 초기 가중치($w_{(1)}$)를 준뒤 결과(y_{out})를 계산한다.
2) 계산 결과와 우리가 원하는 값 사이의 오차를 구한다.
3) 경사하강법을 이용해 바로 앞 가중치를 오차가 작아지는 방향으로 갱신한다.
4) 1~3과정을 더이상 오차가 줄어들지 않을때 까지 반복한다.

여기서 3번에 해당하는 부분을 해 나가기 위해서 (즉 오차를 줄이기 위해서), 경사하강법을 어떻게 적용 해야 할까요? 앞서 배운대로 오차가 적어진다는 의미를, 미분값이 0에 가까워지는 방향으로 나아간다는 말입니다. 즉 '기울기가 0이 되는 방향'으로 나아가야 하지요. 기울기가 0이 되는 방향으로 나아간다는건, 앞서 정한 가중치에서 기울기를 뺐을때 가중치의 변화가 전혀 없는 상태를 말합니다. 따라서 가중치에서 기울기를 빼도 값의 변화가 없을 때 까지 계속해서 이 과정을 반복하는 것이 바로 오차 역전파입니다.

새 가중치는　　현 가중치에서　　"가중치에 대한 기울기"를 빼준 값!

$$W_{(t+1)} = W_t - \frac{\partial \text{오차}}{\partial w}$$

그런데 위 식을 보시면 편미분 기호 ∂(partial, 파셜이라고 읽습니다)가 등장합니다.

이제부터 '미분'이란 말 대신, '편미분(partial derivative)'이란 용어와 방식이 사용되어야 하는데, 이것은 왜 일까요? 잠깐 편미분에 대해 개념을 정리해 보겠습니다.

2. 편미분

미분과 편미분 보두, '미분하라'는 의미에 있어서는 다를바가 없습니다. 그러나, 여러가지 변수가 식 안에 있을때에는 모든 변수를 미분하는 것이 아니라, 우리가 원하는 한가지 변수(예를 들어 가중치)만 미분하고, 그 외에는 모두 상수처럼 취급하라는 뜻이 '편미분'입니다. 앞서 선형회귀를 배울때는 변수가 x하나 뿐이어서 미분하라는 의미에 혼란이 없지만, 변수가 많아질때는 어떤 변수를 미분하는지를 정해야 하므로 편미분이 사용됩니다.

함수 f가 있을때, 이를 x에 관해 편미분하라는 것은, 다음과 같이 표시합니다.

$$\frac{\partial f}{\partial x}$$

편미분의 예를 들어 보겠습니다.

하나 이상의 변수를 갖는 함수 f가 $f(x,y)=x^2+xy+yo2$으로 주어졌다고 가정 할때, $\frac{\partial f}{\partial x}$를 구하려면 어떻게 해야 할까요?

해답은 이렇습니다. 먼저 먼저 미분의 성질, $\frac{d}{dx}(x^n)=nx^{n-1}$에 의해 x^2항은 2x가 됩니다. x에 관해 미분하므로 y는 상수로 취급해야 합니다. 그러면, 앞서 배운 미분 공식 $\frac{d}{dx}(ax)=a$ (a=상수)에 해당됩니다. 따라서 xy가 들어 있는 항은 그냥 y가 됩니다. 또한 $\frac{d}{dx}(a)=0$에 의해 y_{o2}항은 0이 됩니다.

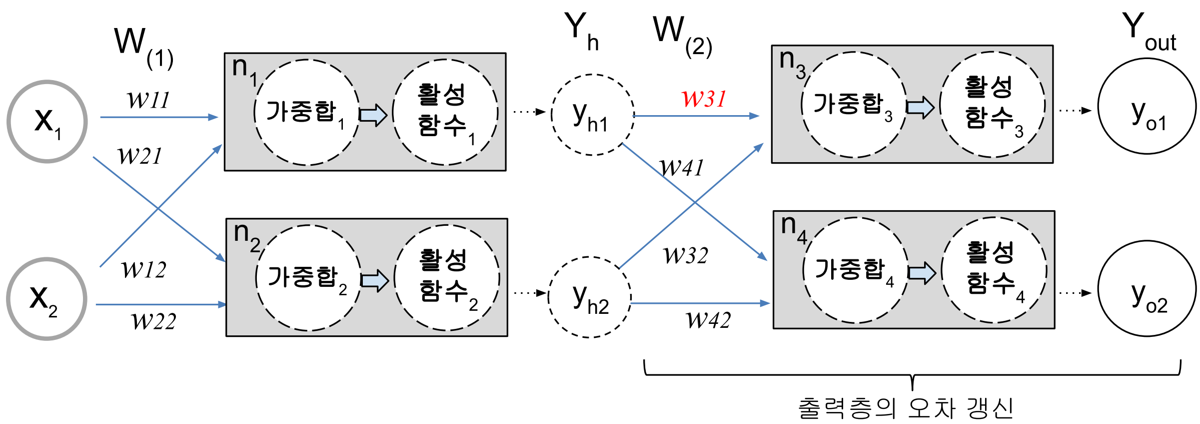
이를 정리하면 아래와 같이 표현할 수 있습니다.

$$f(x,y)=x^2+xy+yo2 \text{ 일때,}$$

$$\frac{\partial f}{\partial x}=2x+y$$

3. 출력층의 오차 갱신

이제 실제로 오차 역전파를 실행해 보겠습니다. 이해를 위해 이번에는 하나의 노드안에서 일어나는 일을 좀더 세부화 시켜서 표시해 보겠습니다. 각 노드는 입력값을 가지고 가중합을 만드는 단계, 가중합을 활성화 함수를 적용해 출력하는 단계로 구분되는데, 각 노드에 이를 표시했습니다. 그리고 출력층의 값은 은닉층의 노드에서 나오는 보이지 않는 출력 값들을 다시한 번 각각의 노드를 거쳐 나오게 되는 과정을 도식으로 표현하면 다음과 같습니다.



[출력층의 오차갱신]

오차역전파는 Y_{out} 값에서 부터 거꾸로 거슬러 올라가며 가중치 $W_{(2)}$와 가중치 $W_{(1)}$을 더이상 업그레이드 되지 않을때 까지 계속해서 업그레이드 시키는 것을 말합니다.

이중, 먼저 $W_{(2)}$의 값 중 하나인 w_{31}을 업데이트 시키는 과정을 알아 보겠습니다.

오차 역전파의 공식을 이용해 w_{31}을 업데이트 하려면 다음 공식을 계산해야 합니다.

$$W_{31}(t+1) = W_{31}t \ - \ \frac{\partial 오차Y_{out}}{\partial w_{31}}$$

W_{31}값은 우리가 이미 알고 있으므로 우리가 실제로 구해야 하는 값은 $\frac{\partial 오차Y_{out}}{\partial w_{31}}$입니다. 이는 오차$Y_{out}$ 를 구하고, 이를 w_{31}에 대하여 편미분하라는 뜻입니다. 이에 먼저 오차Y_{out} 를 구해보겠습니다.

4. 오차 공식

오차Y_{out} 안에는 두개의 출력값이 있습니다(y_{o1}, y_{o2}). 즉 오차Y_{out} = 오차y_{o1} + 오차y_{o2} 입니다. 여기서 오차y_{o1} 과 오차y_{o2} 는 각각 앞서 배운 평균 제곱 오차를 이용해 구합니다. y_{o1}, y_{o2} 의 실제 값을 y_{t1}, y_{t2} 이라고 할때, 다음과 같이 계산됩니다.

$$오차y_{o1} = \frac{1}{2}(y_{t1} - y_{o1})^2$$

$$오차y_{o2} = \frac{1}{2}(y_{t2} - y_{o2})^2$$

그런데 여기서 y_{t1}, y_{t2} 에 해당하는 '실제값'이란 무엇일까요? 바로 데이터에서 얻어낸 y_{o1} 과 y_{o2} 자리의 실제의 값, 즉 우리가 도출해야 하는 정답 값을 말합니다. 이는 계산을 통해 나오는 것이 아니라, 주어진 데이터를 통해 알 수 있는 상수입니다. 우리의 모든 노력은 결국 계산을 통해 나오는 '출력값'이 실제 세상을 통해 알아낸 이 '실제의 값'과 같아 지도록

가중치를 조절해 주는 것이지요. 실제 값은 우리의 목표(target) 이므로 Y_{target}, y_{t1}, y_{t2}라고 표현한 것입니다

이제 오차Y_{out} 는 다음과 같이 구해집니다.

$$오차Y_{out} = \frac{1}{2}(y_{t1} - y_{o1})^2 + \frac{1}{2}(y_{t2} - y_{o2})^2$$

5. 체인 룰

이제, 이 값을 w_{31}에 대하여 편미분 해 보겠습니다.$\frac{\partial 오차Y_{out}}{\partial w_{31}}$의 계산은 합성 함수 미분 공식을 따릅니다. 즉 체인룰(Chain rule)에 의해 다음과 같이 계산할 수 있습니다.

$$\frac{\partial 오차Y_{out}}{\partial w_{31}} = \frac{\partial 오차Y_{out}}{\partial y_{o1}} \cdot \frac{\partial y_{o1}}{\partial 가중합3} \cdot \frac{\partial 가중합3}{\partial w_{31}}$$

체인룰은 연쇄법칙 이라고도 부르며, '합성 함수'를 미분 할 때의 계산 공식입니다.

여기서 합성함수란, 함수 안에 또다른 함수가 들어있는 것을 말합니다. f(x)함수에 들어 있는 'x'값이 또다른 함수 g(x)의 결과일 경우를 말하지요.

예를 들어 우리는 지금 오차Y_{out} 를 미분하려고 합니다. 그런데 오차Y_{out} 는 또다른 식 오차y_{o1} + 오차y_{o2} 의 결과 입니다. 따라서 합성함수가 되었습니다. 그래서 이는 합성함수의 미분이 되는 것이지요.

함성함수는 f(g(x))와 같이 표시합니다. 그리고 이를 미분하면, 안에 있는 g(x)를 x로

대체하여 계산한 값과 g(x)를 미분한 값을 서로 곱해주면 됩니다. 식으로 표현하면 다음과 같습니다.

$$[f(g(x))]' = f'(g(x))g'(x)$$

그리고 이는 아래와 같이 표현할 수도 있지요.

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

여기서 체인 룰, 즉 연쇄 법칙이라고 부르는 이유가 나옵니다. 위 식에서 dg라고 하는 항이 분모와 분자로 고리처럼 연속적으로 이어져 나오기 때문입니다. 우리가 구하려는 합성함수처럼 만일 합성함수식이 3개인 f(g(h(x)))]' 인 경우, 이를 미분하면 다음과 같이 계산 됩니다.

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

6. 체인룰의 계산

이제 체인룰에 의해 주어진 식이 의미하는 바를 하나씩 알아보면서 직접 계산해 보도록 하겠습니다.

$$\frac{\partial \text{오차} Y_{out}}{\partial w_{31}} = \frac{\partial \text{오차} Y_{out}}{\partial y_{o1}} \cdot \frac{\partial y_{o1}}{\partial \text{가중합}3} \cdot \frac{\partial \text{가중합}3}{\partial w_{31}}$$

(i) (ii) (iii)

(i) $\frac{\partial 오차Y_{out}}{\partial y_{o1}}$

먼저 $\frac{\partial 오차Y_{out}}{\partial y_{o1}}$부분입니다. 먼저, 앞서 설명했듯이 오차$Y_{out}$ =오차$_{yo1}$+오차$_{yo2}$ 입니다. 이를 y_{o1}에의해서 편미분하면 오차y_{o2}부분은 상수가 되어 사라지고, 남는 것은 $\frac{\partial 오차y_{o1}}{\partial y_{o1}}$입니다. 여기서 오차$y_{o1}$은 $\frac{1}{2}(y_{t1} - y_{o1})^2$ 이지요. 즉 이제 $\frac{1}{2}(y_{t1} - y_{o1})^2$ 를 y_{o1}으로 편미분 하면 그 결과값은 y_{o1}-y_{t1}이 됩니다.

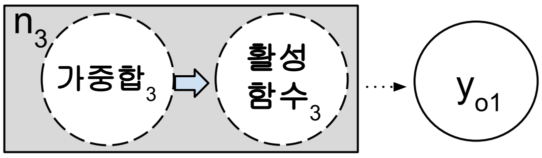
> y_{o1}-y_{t1} 로 편미분 되는 과정을 유도하면 다음과 같습니다.
>
> $$\frac{\partial 오차Y_{out}}{\partial y_{o1}} = \frac{\partial(오차y_{o1}+오차y_{o2})}{\partial y_{o1}}$$
>
> $$= \frac{\partial \frac{1}{2}(y_{t1}-y_{o1})^2}{\partial y_{o1}} + \frac{\partial \frac{1}{2}(y_{t2}-y_{o2})^2}{\partial y_{o1}}$$
>
> $$= [\frac{1}{2}(y_{t1} - y_{o1})^2]' + (y_{t1} - y_{o1})'$$
>
> $= (y_{t1} - y_{o1})(-1)$
>
> $= y_{o1}-y_{t1}$

따라서 다음과 같이 정리되었습니다.

$$\frac{\partial 오차Yout}{\partial yo1} = y_{o1} - y_{t1} \quad \text{---- (1)}$$

(ii) $\frac{\partial y_{o1}}{\partial 가중합_3}$

다음으로$\frac{\partial y_{o1}}{\partial 가중합_3}$ 부분을 보겠습니다. 이를 위해 위 그림 [출력층의 오차갱신]의 일부를 한번 더 보시겠습니다.



여기서 가중합$_3$이 y_{o1}으로 바뀌는 과정에는 활성함수(활성화 함수)$_3$ 을 거치는 것을 알수 있습니다. 가중합$_3$이 활성함수$_3$을 통해 y_{o1}이 되는 것입니다. 그러면, 가중합$_3$을 y_{o1}에 대해 미분하라는 것은 y_{o1}을 배출한 활성함수$_3$을 미분하라는 뜻이 됩니다.

$$\frac{\partial y_{o1}}{\partial 가중합_3} = 활성함수_3의\ 미분$$

활성화 함수로 앞서 배운 여러가지가 있지만, 우리는 그 중 시그모이드 함수를 써보도록 하겠습니다. 그러면 이제 시그모이드 함수를 미분하는 방법을 알아볼 차례입니다.

함수 σ(x) 를 시그모이드 함수 $\frac{1}{1+e^{-x}}$로 정의 할 때, 이를 미분한 $\frac{d\sigma(x)}{dx}$값은, 다음과 같습니다.

$$\frac{d\sigma(x)}{dx} = \sigma(x)\cdot(1-\sigma(x))$$

시그모이드 함수의 미분은 시그모이드 값과 그 값을 1에서 빼준 값을 곱하면 되는 것이지요.

시그모이드 함수를 미분 과정을 유도하면 다음과 같습니다. 궁금하신 분들은 아래 증명을 참조하시기 바랍니다.

$\sigma(x) = \frac{1}{1+e^{-x}}$일때,

$$
\begin{aligned}
\frac{d}{dx}\sigma(x) &= \frac{d}{dx}\left[\frac{1}{1+e^{-x}}\right] \\
&= \frac{d}{dx}\left(1+\mathrm{e}^{-x}\right)^{-1} \\
&= -(1+e^{-x})^{-2}(-e^{-x}) \\
&= \frac{e^{-x}}{(1+e^{-x})^2} \\
&= \frac{1}{1+e^{-x}} \cdot \frac{e^{-x}}{1+e^{-x}} \\
&= \frac{1}{1+e^{-x}} \cdot \frac{(1+e^{-x})-1}{1+e^{-x}} \\
&= \frac{1}{1+e^{-x}} \cdot \left(1-\frac{1}{1+e^{-x}}\right) \\
&= \sigma(x) \cdot (1-\sigma(x))
\end{aligned}
$$

연쇄법칙 적용

1. $f(x)=x^a$ (a=자연수)일때, 미분값은 ax^{a-1}
2. e^{-x} 의 미분값은 $-e^x$

증명:

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}x}\left[\mathrm{e}^{-x}\right] \\
&= \mathrm{e}^{-x} \cdot \frac{\mathrm{d}}{\mathrm{d}x}[-x] \\
&= \left(-\frac{\mathrm{d}}{\mathrm{d}x}[x]\right)\mathrm{e}^{-x} \\
&= -1\mathrm{e}^{-x} \\
&= -\mathrm{e}^{-x}
\end{aligned}
$$

여기서 활성함수$_3$의 값은 y_{o1}입니다. 따라서 활성함수$_3$의 미분은 다음과 같이 구할 수 있습니다.

$$활성함수_3의\ 미분 = y_{o1} \cdot (1-y_{o1})$$

이제 주어진 $\frac{\partial y_{o1}}{\partial 가중합_3}$ 식을 정리하면 다음과 같습니다.

$$\text{(ii)} \quad \frac{\partial y_{o1}}{\partial 가중합_3} = y_{o1} \cdot (1-y_{o1}) \quad \text{--- (2)}$$

(iii) $\frac{\partial 가중합_3}{\partial w_{31}}$

이제 마지막 남은 $\frac{\partial 가중합_3}{\partial w_{31}}$항 입니다.

여기서, 가중합$_3$은 n1,n2노드로 부터 전달된 Y_h값과 $W_{(1)}$의 값을 통해 만들어 집니다.

$$가중합_3 = w_{31}y_{h1} + w_{41}y_{h2} + 1(바이어스)$$

그런데 앞장에서 자세히 다루었던 바이어스가 여기서는 1로 대체되어 사용됩니다. 신경망에서 바이어스를 항상 1로 설정해 놓습니다. 왜 그럴까요? 바이어스의 역할은 앞서 배웠듯이 그래프를 좌표상에서 좌, 우로 움직여 주는 역할을 하지요. 활성화 함수로 사용되는 시그모이드 함수에 있어서 가장 안정된 예측이 가능하게 해 주는 바이어스 값은 바로 1입니다. 따라서 바이어스 값을 따로 계산하지 않고 1로 처리하여 연산의 속도를 더해 줍니다.

이를 W_{31}에 관하여 편미분하므로, w_{31}과 관계없는 $w_{41}y_{h2}$ 과 바이어스 항은 모두 상수처리되어 사라집니다. 따라서 남아 있는 $w_{31}y_{h1}$ 항을 미분하면 다음과 같이 정리 됩니다.

$$\text{(iii)} \quad \frac{\partial 가중합_3}{\partial w_{31}} = y_{h1} \quad \text{--- (3)}$$

이제 (1),(2),(3)을 모두 정리하면 다음과 같이 간단하게 정리됩니다.

$$\frac{\partial 오차Y_{out}}{\partial w_{31}} = \frac{\partial 오차Y_{out}}{\partial y_{o1}} \cdot \frac{\partial y_{o1}}{\partial 가중합_3} \cdot \frac{\partial 가중합_3}{\partial w_{31}}$$

$$=(y_{o1}- y_{t1}).y_{o1}(1-y_{o1})\ .y_{h1}$$

7. 가중치 업데이트

앞서구한 값을 w_{31}에서 빼주면 새로운 w_{31}값을 구하게 되는 것입니다. 따라서 출력층의 가중치를 갱신하는 방법을 다음과 같이 정리할 수 있습니다.

$$W_{31}(t+1) = W_{31}t\ - (y_{o1}- y_{t1}).y_{o1}(1-y_{o1})\ \cdot y_{h1}$$

이제 다음으로 넘어가기 전, 위 식에서 y_{h1}앞에 나오는 부분의 형태를 잘 보시기 바랍니다.

$$(y_{o1}- y_{t1}).y_{o1}(1-y_{o1})$$

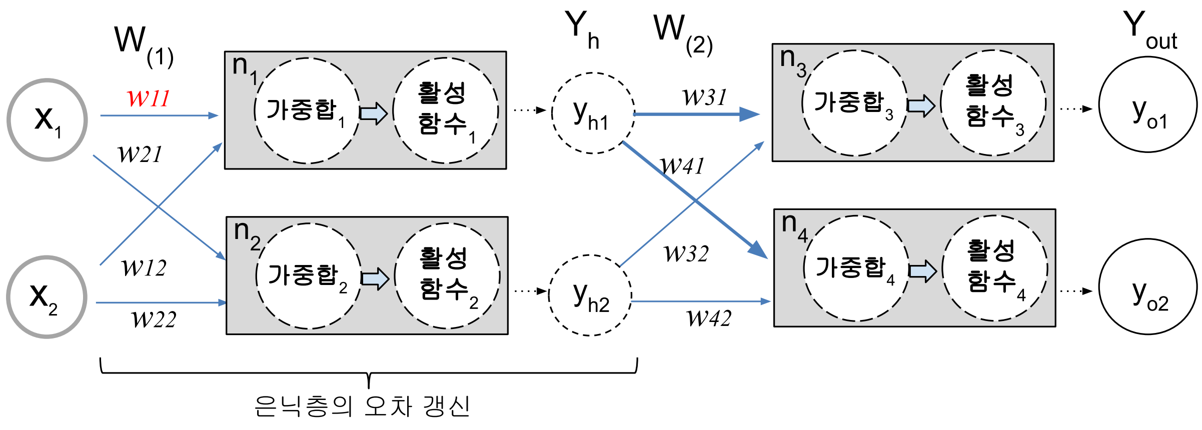
다음 장에서 배우겠지만, 지금 이 형태는 다음 오차의 갱신 때에도 반복해서 나타납니다. 따라서 이 식을 한번 구해 놓으면, 이후는 이 식을 그대로 사용하여 오차를 구할 수 있습니다. 이를 node3의 델타(delta)식 이라고 합니다.

이 델타식을 δy라고 하면, 우리가 해내야 하는 오차의 갱신은 다음과 같은 식으로도 구할 수 있습니다.

$$W_{31}(t+1) = W_{31}t\ - \delta y \cdot y_{h1}$$

8. 은닉층의 오차 갱신

이제 출력층을 거쳐 은닉층의 오차가 갱신되는 과정을 살펴보겠습니다. 마찬가지로 은닉층의 오차 $W_{(2)}$ 중 하나인, w_{11}의 값을 업데이트 하는 방법을 설명하겠습니다.



[은닉층의 오차 갱신]

마찬가지로 가중치에 기울기를 뺀 값을 구해야 합니다. 이때, 우리가 구하려는 값이 w_{11}이므로 다음과 같이 계산합니다.

$$W_{11}(t+1) = W_{11}t - \frac{\partial 오차Y_{out}}{\partial W_{11}}$$

여기서 $\frac{\partial 오차Y_h}{\partial w_{11}}$가 아니라 $\frac{\partial 오차Y_{out}}{\partial w_{11}}$ 이라는 점을 주목하시기 바랍니다. 그 이유는 Y_h 은닉층 안에 위치해 있으므로 겉으로 드러나지 않습니다. 따라서 그 값을 알수 없습니다. 우리가 알수 있는 출력값은 Y_{out}뿐이므로, 은닉층의 오차 갱신을 위한 기울기를 구할때에도 Y_{out}에서부터 출발해야 합니다.

이제 앞서 계산했던 바와 마찬가지로 기울기에 해당하는 $\frac{\partial 오차Y_{out}}{\partial w_{11}}$ 를 구해 보겠습니다. 체인룰을 적용하여, 다음과 같이 계산됩니다.

$$\frac{\partial 오차Y_{out}}{\partial w_{11}} = \frac{\partial 오차Y_{out}}{\partial y_{h1}} \cdot \frac{\partial y_{h1}}{\partial 가중합_1} \cdot \frac{\partial 가중합_1}{\partial w_{11}}$$

(i) (ii) (iii)

여기서 (ii) $\frac{\partial y_{h1}}{\partial 가중합_1}$과 (iii) $\frac{\partial 가중합_1}{\partial w_{11}}$ 항은 이전과 동일한 방법으로 계산됩니다. 따라서 다음과 같이 바꿔줄 수 있습니다.

$$\frac{\partial y_{h1}}{\partial 가중합_1} \cdot \frac{\partial 가중합_1}{\partial w_{11}} = y_{h1}(1-y_{h1}) \cdot x_1$$

9. 은닉층의 오차계산

그런데, 은닉층에서의 (i) $\frac{\partial 오차Y_{out}}{\partial y_{h1}}$항은 계산이 조금 다릅니다. 오차$Y_{out}$ 안에는 오차y_{o1}과 오차y_{o2}가 포함되어 있습니다. 이전에는 오차Y_{out}를 y_{o1}에 의해서 편미분할때, y_{o1} 와 관계없는 항인 오차y_{o2}는 상수가 되어 사라졌습니다. 따라서 남는 것은 $\frac{\partial 오차Y_{o1}}{\partial y_{o1}}$ 뿐이었습니다. 하지만, 이번에는 y_{h1}에 대해 미분해야 합니다. [은닉층의 오차갱신] 그림에서 굵은 화살표로 표시된 부분을 보면 알수 있듯이, y_{h1}은 오차$_{yo1}$과 오차$_{yo2}$ 의 형성에 모두 관계가 있습니다. 따라서 $\frac{\partial 오차Y_{o1}}{\partial y_{h1}}$과$\frac{\partial 오차Y_{o2}}{\partial y_{h1}}$이 모두 계산되어야 하므로 계산이 다음과 같이 조금 복잡해 집니다.

$$\text{(i)} \quad \frac{\partial 오차Y_{out}}{\partial h_1} = \frac{\partial(오차y_{o1} + 오차y_{o2})}{\partial y_{h1}} = \frac{\partial 오차y_{o1}}{\partial y_{h1}} + \frac{\partial 오차y_{o2}}{\partial y_{h1}}$$

| |
(a) (b)

(a) $\dfrac{\partial 오차y_{o1}}{\partial y_{h1}}$

먼저 (a) $\frac{\partial 오차y_{o1}}{\partial y_{h1}}$를 보시겠습니다. 체인룰에 의해 다음과 같이 바뀝니다.

$$\frac{\partial 오차y_{o1}}{\partial y_{h1}} = \frac{\partial 오차y_{o1}}{\partial 가중합_3} \cdot \frac{\partial 가중합_3}{\partial y_{h1}}$$

이 중 $\frac{\partial 오차y_{o1}}{\partial 가중합_3}$부분을 다시 미분하면 역시 체인룰에 의해 다음과 같이 바뀝니다.

$$\frac{\partial 오차y_{o1}}{\partial y_{o1}} \cdot \frac{\partial y_{o1}}{\partial 가중합_3}$$

이중 먼저 $\frac{\partial 오차y_{o1}}{\partial y_{o1}}$에서 오차$y_{o1}$은 $\frac{1}{2}(y_{t1} - y_{o1})^2$이므로 y_{o1}으로 편미분하면 $y_{o1} - y_{t1}$ 이 됩니다.

그리고 $\frac{\partial y_{o1}}{\partial 가중합_3}$는 앞서 설명한 대로 시그모이드 함수의 미분입니다. 따라서 $y_{o1} \cdot (1-y_{o1})$로 계산됩니다.

이제 나머지 $\frac{\partial 가중합_3}{\partial y_{h1}}$를 미분하면 y_{o1}가 남습니다.

모두 정리하면 다음과 같습니다.

$$\frac{\partial 오차yo1}{\partial y_{h1}} = (y_{o1} - y_{t1}).y_{o1}(1-y_{o1}) .y_{o1}$$

그런데 여기서 $(y_{o1} - y_{t1}).y_{o1}(1-y_{o1})$부분이 눈에 익으신지요. 앞서 기억해 두었던 델타식(δy)의 형식 입니다. 지금 우리는 y_{o1}을 구해야 하므로 델타식을 δy_{o1}라고 할때, 위값은 아래와 같이 간단하게 표시할 수도 있습니다.

$$(a)\frac{\partial 오차y_{o1}}{\partial y_{h1}} = \delta y_{o1}.y_{o1} \quad ---(1)$$

(b) $\frac{\partial 오차 y_{o2}}{\partial y_{h1}}$

이제 (b) $\frac{\partial 오차 y_{o2}}{\partial y_{h1}}$부분을 보시겠습니다. 역시 체인룰에 의해 다음과 같이 변형됩니다.

$$\frac{\partial 오차 y_{o2}}{\partial y_{h1}} = \frac{\partial 오차 y_{o2}}{\partial 가중합_4} \cdot \frac{\partial 가중합_4}{\partial y_{h1}}$$

이 중 $\frac{\partial 오차 y_{o2}}{\partial 가중합4}$부분은 체인룰에 의해 $\frac{\partial 오차 y_{o2}}{\partial y_{o2}} \cdot \frac{\partial y_{o2}}{\partial 가중합_4}$로 바뀝니다.

앞서 (a)식의 풀이에서 설명한 방식과 똑같이 적용되므로 답을 바로 구해보면, 다음과 같습니다.

$$\frac{\partial 오차 y_{o2}}{\partial y_{o2}} = (y_{o2} - y_{t2}).y_{o2}(1-y_{o2})\ .y_{o2}$$

여기서 $(y_{o2} - y_{t2}).y_{o2}(1-y_{o2})$ 부분은 역시 델타식 형식이지요. 이 델타식을 δy_{o1}라고 할때, 주어진 식은 다음과 같이 표시할 수 있습니다.

$$(b) \frac{\partial 오차 y_{o2}}{\partial y_{h1}} = \delta y_{o2}.y_{o2} \quad \text{---(2)}$$

이제 (1), (2)를 종합하면 다음과 같습니다.

$$(i)\ \frac{\partial 오차 y_{o1}}{\partial y_{h1}} + \frac{\partial 오차 y_{o2}}{\partial y_{h1}} = \delta y_{o1}.y_{o1} + \delta y_{o2}.y_{o2}$$

이제 앞서 구한 (i), (ii), (iii)의 모든 값을 다음과 정리해 볼 수 있습니다.

$$\frac{\partial 오차Y_{out}}{\partial y_{h1}} = \frac{\partial 오차Y_{out}}{\partial y_{h1}} \cdot \frac{\partial y_{h1}}{\partial 가중합_1} \cdot \frac{\partial 가중합_1}{\partial w_{11}}$$

$$= (\delta y_{o1}.y_{o1} + \delta y_{o2}.y_{o2})y_{h1}(1-y_{h1}) .x_1$$

10. 델타식

이제 출력층과 은닉층의 갱신을 위해 도출된 두개의 식을 비교해 보겠습니다.

$$출력층의\ 오차\ 갱신 = \underbrace{(y_{o1} - y_{t1}).y_{o1}(1-y_{o1})}_{a} .y_{h1}$$

$$은닉층의\ 오차\ 갱신 = \underbrace{(\delta y_{o1}.y_{o1} + \delta y_{o2}.y_{o2})y_{h1}(1-y_{h1})}_{b} .x_1$$

여기서 노란색으로 표시된 두 부분을 비교해 보겠습니다. a의 $(y_{o1} - y_{t1})$가 b 에서 $(\delta y_{o1}.y_{o1} + \delta y_{o2}.y_{o2})$ 로 바뀌었지만, 나머지 부분은 *out*(1-*out*)의 형태를 띄고 있습니다.

여기서 $(y_{o1} - y_{t1})$는 오차의 값입니다. 하지만, 은닉층에서는 이렇게 오차를 계산할 수 없습니다. 은닉층에서 일어나는 일은 우리가 겉에서 알수가 없기 때문에 출력층으로 부터 y_o값을 가져와 계산해야 하지요. 그러므로 위 식과 같이 $(\delta y_{o1}.y_{o1} + \delta y_{o2}.y_{o2})$ 으로 형태가 복잡해 졌을 뿐, 결국 오차를 나타냅니다. 따라서 두 식 모두 "오차.*out*(1-*out*)"의 형태, 즉 델타식의 형태로 단순화 할 수 있습니다.

델타식이 중요한 이유는 이렇게 한층을 거슬러 올라갈때 마다 같은 형태가 계속 나타나기 때문입니다. 따라서 델타식을 파악하고 나면 이를 코딩으로 설계하는 것도 어렵지 않습니다.

은닉층의 델타식이므로 이것을 δh라고 할때, 은닉층의 가중치 갱신을 식으로 표현하면 다음과 같습니다.

$$W_{11}(t+1) = W_{11}t - \delta h \cdot x_1$$

<오타 제보> taehjo@gmail.com

<참고문헌> Gorman, R. Paul, and Terrence J. Sejnowski. "Analysis of Hidden Units in a Layered Network Trained to Classify Sonar Targets." Neural Networks 1.1 (1988): 75-89.