심화 학습

심화 학습 1 오차 역전파의 계산법

1 출력층의 오차 업데이트

2 오차 공식

3 체인 룰

4 체인 룰 계산하기

5 가중치 수정하기

6 은닉층의 오차 수정하기

7 은닉층의 오차 계산법

8 델타식

오차 역전파의 계산법

- **오차 역전파의 계산법**
  - 케라스, 텐서플로 같은 딥러닝 라이브러리를 적극적으로 활용하면 연구 및 산업 현장에서 만나는 대부분 프로젝트를 해낼 수 있음
  - 더 나은 결과를 얻고 싶거나 딥러닝 알고리즘 자체를 공부한다면 신경망의 핵심인 오차 역전파의 계산법을 완전히 이해하고 이를 통해 신경망을 더욱 깊이 통찰할 수 있어야 함
  - ‘심화 학습 1’에서는 오차 역전파의 개념을 설명하고 그 계산 과정을 살펴보자

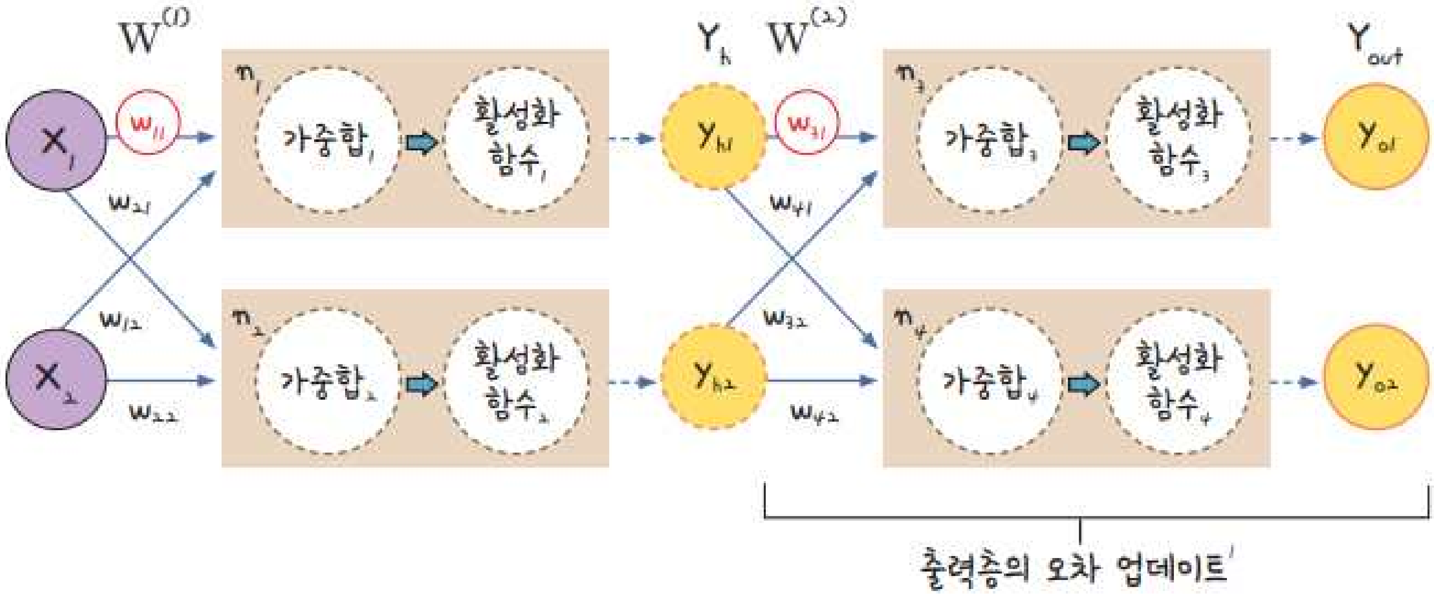
1 출력층의 오차 업데이트

1 출력층의 오차 업데이트

- **출력층의 오차 업데이트**
  - 이제 실제로 오차 역전파를 실행해 보자
  - 이해를 돕고자 노드 하나 안에서 일어나는 일을 좀 더 세분화해서 표시
  - 각 노드 내부에서는 입력 값을 이용해 가중합을 만드는 단계와 이 가중합을 활성화 함수를 적용해 출력하는 단계로 구분
  - 이 두 단계를 각 노드 내부에 표시하고 각 가중치(w) 값과 은닉층의 출력 값(Y_h)을 포함해 표현하면 다음 그림과 같음

1 출력층의 오차 업데이트

▼ 그림 1 | 출력층의 오차 업데이트



1 출력층의 오차 업데이트

- **출력층의 오차 업데이트**
  - 오차 역전파는 Y_{out}값에서 거꾸로 거슬러 올라가며 가중치 $W^{(2)}$와 가중치 $W^{(1)}$이 더는 업데이트되지 않을 때까지 반복해 계산하는 것
  - 먼저 $W^{(2)}$의 값 중 하나인 w_{31}을 업데이트하는 과정을 알아보자
  - 오차 역전파의 공식을 이용해 w_{31}을 업데이트하려면 다음 공식으로 계산

$$w_{31}(t+1) = w_{31}t - \frac{\partial \text{오차} Y_{out}}{\partial w_{31}}$$

1 출력층의 오차 업데이트

- **출력층의 오차 업데이트**
  - 여기서 t는 한 단계 앞, t + 1은 현재 단계의 계산을 의미
  - $w_{31}t$는 한 단계 앞에서 이미 계산된 값을 의미하므로 여기서는 구할 필요가 없음
  - 우리가 실제로 구해야 하는 값은 $\frac{\partial 오차 Y_{out}}{\partial w_{31}}$
  - 이는 오차 Y_{out}을 구하고 이를 w_{31}에 대해 편미분하라는 의미
  - 먼저 오차 Y_{out}을 구해 보자

2 오차 공식

2 오차 공식

- **오차 공식**
  - 오차 Y_{out} 안에는 두 개(y_{o1}, y_{o2})의 출력 값이 있음
  - 즉, 오차 Y_{out} = 오차 y_{o1} + 오차 y_{o2}
  - 여기서 오차 y_{o1}과 오차 y_{o2}는 각각 앞서 배운 평균 제곱 오차를 이용해 구함
  - y_{o1}, y_{o2}의 실제 값을 y_{t1}, y_{t2}라고 할 때, 다음과 같이 계산

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

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

2 오차 공식

오차 공식

- 여기서 y_{t1}, y_{t2}에 해당하는 '실제 값'이란 무엇일까?
- 바로 데이터에서 얻어 낸 y_{o1}과 y_{o2}자리의 실제 값, 즉 도출해야 하는 정답 값을 의미
- 이는 계산해서 나오는 것이 아니라, 주어진 데이터를 통해 알 수 있는 상수
- 결국 우리가 해야 할 일은 계산해서 나오는 '출력 값'이 실제 세상을 통해 알아낸 '실제 값'과 같아지도록 가중치를 조절해 주는 것
- 실제 값은 우리의 목표(target)이므로 Y_{target}, y_{t1}, y_{t2}라고 표현한 것
- y_{t1}, y_{t2}에서의 t는 target, 즉 우리가 구해야 할 목표를 의미

2 오차 공식

- **오차 공식**
  - 이제 오차 Y_{out}은 다음과 같이 구할 수 있음

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

3 체인 룰

3 체인 룰

- **체인 룰**
  - 이제 이 값을 w_{31}에 대해 편미분해 보자
  - $\frac{\partial 오차 Y_{out}}{\partial w_{31}}$ 의 계산은 합성 함수 미분 공식을 따름
  - 즉, 체인 룰(chain rule)에 의해 다음과 같이 계산할 수 있음

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

3 체인 룰

- 체인 룰
  - 체인 룰은 연쇄 법칙이라고도 하며, '합성 함수'를 미분할 때의 계산 공식
  - 여기서 합성 함수란 함수 안에 또 다른 함수가 들어 있는 것을 의미
  - f(x) 함수에 들어 있는 x 값이 또 다른 함수 g(x)의 결과일 때를 의미
  - 예를 들어 우리는 지금 오차 Y_{out}을 미분하려고 함
  - 오차 Y_{out}은 또 다른 식의 오차 y_{o1} + 오차 y_{o2}의 결과
  - 합성 함수가 됨
  - 이는 합성 함수의 미분이 되는 것

3 체인 룰

- 체인 룰
  - 합성 함수는 f(g(x))처럼 표시
  - 이를 미분하면 안에 있는 g(x)를 x로 대체해 계산한 값과 g(x)를 미분한 값을 서로 곱해 주면 됨
  - 식으로 표현하면 다음과 같음

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

  - 이는 다음과 같이 표현할 수도 있음

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

3 체인 룰

- 체인 룰
  - 여기서 체인 룰, 즉 연쇄 법칙이라고 하는 이유가 나옴
  - 앞의 식에서 dg라고 하는 항이 분모와 분자로 고리처럼 연속적으로 이어져 나오기 때문임
  - 우리가 구하려는 합성 함수처럼 만일 합성 함수식이 세 개라면, 즉 f(g(h(x)))′일 때 이를 미분하면 다음과 같음

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

4 체인 룰 계산하기

4 체인 룰 계산하기

- **체인 룰 계산하기**
  - 이제 체인 룰을 사용해서 주어진 식이 의미하는 것을 하나씩 알아보면서 직접 계산해 보자

$$\frac{\partial 오차Y_{out}}{\partial w_{31}} = \underbrace{\frac{\partial 오차Y_{out}}{\partial y_{o1}}}_{①} \cdot \underbrace{\frac{\partial y_{o1}}{\partial 가중합_3}}_{②} \cdot \underbrace{\frac{\partial 가중합_3}{\partial w_{31}}}_{③}$$

① $\frac{\partial 오차Y_{out}}{\partial y_{o1}}$

4 체인 룰 계산하기

- **체인 룰 계산하기**
  - 앞서 6.3절에서 설명했듯이 오차 Y_{out}= 오차 y_{o1} + 오차 y_{o2}
  - 이를 yo1에 의해 편미분하면 y_{o1}과 관계없는 y_{o2} 부분은 상수가 되어 사라지고, 남는 것은

$$\frac{\partial 오차 y_{o1}}{\partial y_{o1}}$$

  - 여기서 오차 y_{o1}은 $\frac{1}{2}(y_{t1} - y_{o1})^2$
  - 이제 $\frac{1}{2}(y_{t1} - y_{o1})^2$ 을 y_{o1}로 편미분하면 결괏값은 y_{o1} - y_{t1}이 됨

4 체인 룰 계산하기

체인 룰 계산하기

편미분 과정 유도

- y_{o1} - y_{t1}로 편미분되는 과정을 유도하면 다음과 같음

$$
\begin{aligned}
\frac{\partial 오차 Y_{out}}{\partial y_{o1}} &= \frac{\partial 오차 y_{o1} + \partial 오차\ y_{o2}}{\partial y_{o1}} \\
&= \frac{\partial \frac{1}{2}(y_{t1}-y_{o1})^2}{\partial y_{o1}} + \frac{\partial \frac{1}{2}(y_{t2}-y_{o2})^2}{\partial y_{o1}} \\
&= \frac{1}{2}(y_{t1}-y_{o1})^2 + (y_{t2}-y_{o2})' \\
&= (y_{t1}-y_{o1})(-1) \\
&= y_{o1}-y_{t1}
\end{aligned}
$$

4 체인 룰 계산하기

- **체인 룰 계산하기**
  - 다음과 같이 정리할 수 있음

$$\frac{\partial \text{오차}Y_{\text{out}}}{\partial y_{o1}} = y_{o1} - y_{t1} \cdots ⓐ$$

$$② \ \frac{\partial y_{o1}}{\partial \text{가중합}_3}$$

4 체인 룰 계산하기

- 체인 룰 계산하기
  - 이 부분을 보기 앞서 다음 그림을 한 번 더 보자

▼ 그림 2 | 가중합과 활성화 함수

4 체인 룰 계산하기

- 체인 룰 계산하기
  - 가중합$_3$이 y_{o1}로 바뀌는 과정에는 활성화 함수$_3$을 거치는 것을 알 수 있음
  - 가중합$_3$이 활성화 함수$_3$을 통해 y_{o1}이 됨
  - y_{o1}을 가중합$_3$에 대해 미분하라는 것은 y_{o1}을 배출한 활성화 함수$_3$을 미분하라는 의미가 됨

$$\frac{\partial y_{o1}}{\partial \text{가중합}_3} = \text{활성화 함수}_3\text{의 미분}$$

4 체인 룰 계산하기

- **체인 룰 계산하기**
  - 활성화 함수에는 여러 가지가 있지만, 그중에서 시그모이드 함수를 사용
  - 이제 시그모이드 함수를 미분하는 방법을 알아볼 차례
  - 함수 $\sigma(x)$ 를 시그모이드 함수 $\frac{1}{1+e^{-x}}$ 로 정의할 때 이를 미분한 $\frac{d\sigma(x)}{dx}$ 값은 다음과 같음

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

  - 다시 말해 시그모이드 함수의 미분은 시그모이드 값과 그 값을 1에서 뺀 값을 곱하면 됨

4 체인 룰 계산하기

- **체인 룰 계산하기**

  시그모이드 함수를 미분하는 과정

  - 다음은 시그모이드 함수의 미분을 유도하는 과정
  - 궁금하다면 다음 증명을 참고

4 체인 룰 계산하기

● 체인 룰 계산하기

$\sigma(x) = \frac{1}{1+e^{-x}}$ 일 때

$$\frac{d}{dx}\sigma(x) = \frac{d}{dx}\left[\frac{1}{1+e^{-x}}\right]$$
$$= \frac{d}{dx}(1+e^{-x})^{-1}$$
$$= -(1+e^{-x})^{-2}(-e^{-x})$$

← 연쇄 법칙 적용

1. $f(x) = x^a$ (a = 자연수)일 때, 미분 값은 ax^{a-1}
2. e^{-x}의 미분 값은 $-e^{-x}$

증명: $\frac{d}{dx}[e^{-x}]$
$$= e^{-x} \cdot \frac{d}{dx}[-x]$$
$$= (-\frac{d}{dx}[x])e^{-x}$$
$$= -1e^{-x}$$
$$= -e^{-x}$$

$$= \frac{e^{-x}}{(1+e^{-x})^2}$$
$$= \frac{1}{1+e^{-x}} \cdot \frac{e^{-x}}{1+e^{-x}}$$
$$= \frac{1}{1+e^{-x}} \cdot \frac{(1+e^{-x})-1}{1+e^{-x}}$$
$$= \frac{1}{1+e^{-x}} \cdot (1-\frac{1}{1+e^{-x}})$$
$$= \sigma(x)\cdot(1-\sigma(x))$$

4 체인 룰 계산하기

- **체인 룰 계산하기**
  - 여기서 활성화 함수$_3$의 값은 y_{o1}
  - 활성화 함수$_3$의 미분은 다음과 같이 구할 수 있음

$$\text{활성화 함수}_3\text{의 미분} = y_{o1} \cdot (1 - y_{o1})$$

4 체인 룰 계산하기

- **체인 룰 계산하기**
  - 이제 주어진 $\frac{\partial y_{o1}}{\partial 가중합_3}$ 식을 정리하면 다음과 같음

$$\frac{\partial y_{o1}}{\partial 가중합_3} = y_{o1} \cdot (1 - y_{o1}) \cdots \text{ⓑ}$$

3 $\frac{\partial 가중합_3}{\partial w_{31}}$

4 체인 룰 계산하기

- **체인 룰 계산하기**
  - 여기서 가중합$_3$은 n_1과 n_2 노드로부터 전달된 y_h값과 $w_{(2)}$ 값을 통해 만들어짐

$$\text{가중합}_3 = w_{31} y_{h1} + w_{31} y_{h2} + 1\,(\text{바이어스})$$

4 체인 룰 계산하기

- **체인 룰 계산하기**
  - 책에서 다루었던 바이어스가 여기서는 1로 대체되어 사용
  - 신경망에서는 바이어스를 항상 1로 설정해 놓는데 왜 그럴까?
  - 바이어스는 그래프를 좌표에서 좌우로 움직이는 역할을 하는데, 활성화 함수로 사용되는 시그모이드 함수가 가장 안정된 예측을 하게 하는 바이어스 값이 1이기 때문임
  - 바이어스 값을 따로 계산하지 않고 1로 처리해 연산 속도를 높임

4 체인 룰 계산하기

- 체인 룰 계산하기
  - w_{31}에 대해 편미분하므로 w_{31}과 관계없는 $w_{32}y_{h2}$와 바이어스 항은 모두 상수로 처리되어 사라짐
  - 남은 $w_{31}y_{h1}$ 항을 미분하면 다음과 같이 정리

$$\frac{\partial 가중합_3}{\partial w_{31}} = y_{h1} \cdots \text{C}$$

4 체인 룰 계산하기

- **체인 룰 계산하기**
  - 이제 a b c 를 이용해 주어진 식을 다시 한 번 정리하면 다음과 같음

$$\frac{\partial \text{오차}Y_{\text{out}}}{\partial w_{31}} = \frac{\partial \text{오차}Y_{\text{out}}}{\partial y_{o1}} \cdot \frac{\partial y_{o1}}{\partial \text{가중합}} \cdot \frac{\partial \text{가중합}_3}{\partial w_{31}}$$

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

5 가중치 수정하기

5 가중치 수정하기

- **가중치 수정하기**
  - 앞서 구한 값을 w_{31}에서 빼 주면 새로운 w_{31} 값을 구할 수 있음
  - 출력층의 가중치를 업데이트하는 방법을 다음과 같이 정리할 수 있음

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

  - 이제 다음으로 넘어가기 전 앞의 식에서 y_{h1} 앞에 나오는 부분의 형태를 잘 기억해 두자

$$(y_{o1} - y_{t1}) \cdot y_{o1}(1 - y_{o1})$$

5 가중치 수정하기

- **가중치 수정하기**
  - 다음 장에서 배우겠지만, 이 형태는 다음 오차 업데이트 때도 반복해서 나타남
  - 이 식을 한 번 구해 놓으면 이후는 그대로 사용해서 오차를 구할 수 있음
  - 이를 node3의 **델타**(delta)**식**이라고 함

5 가중치 수정하기

- **가중치 수정하기**
  - 이 델타식을 δy 라고 하면 우리가 해내야 하는 오차의 업데이트는 다음 식으로도 구할 수 있음

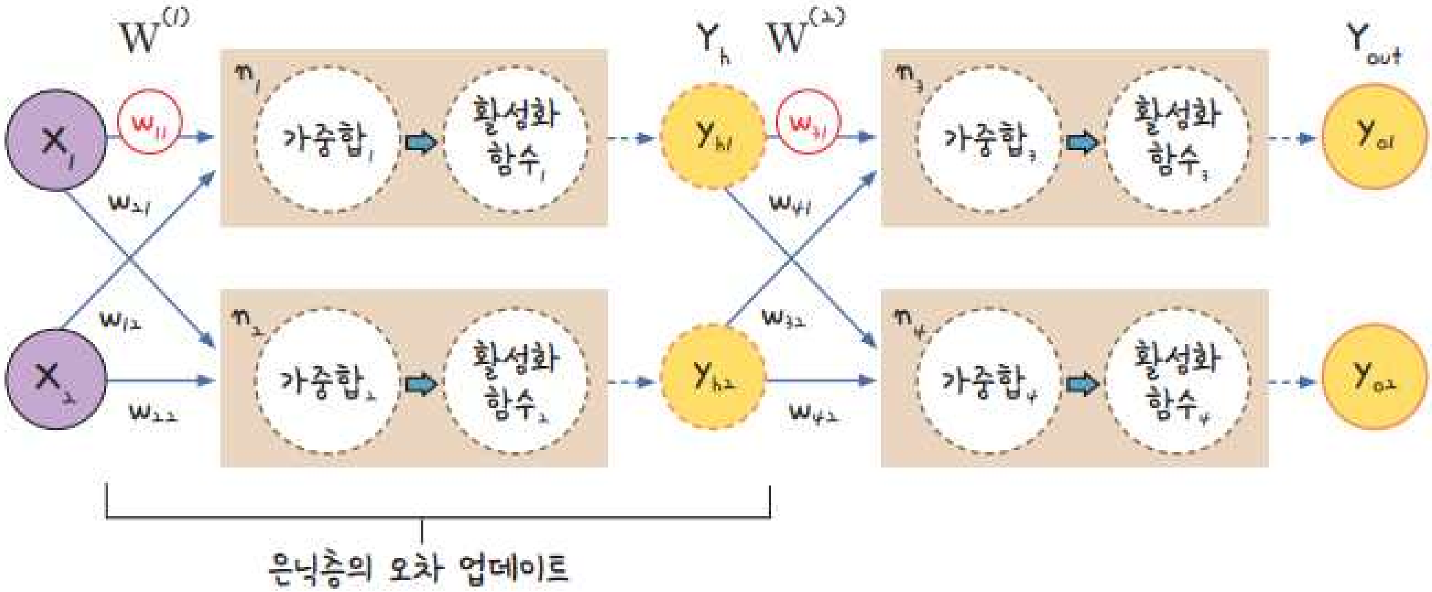
$$w_{31}(t+1) = w_{31}t - \delta y \cdot y_{h1}$$

6 은닉층의 오차 수정하기

6 은닉층의 오차 수정하기

- 은닉층의 오차 수정하기
  - 이제 출력층을 거쳐 은닉층의 오차가 업데이트되는 과정을 살펴보자
  - 마찬가지로 은닉층의 오차 $W^{(1)}$ 중 하나인 w_{11} 값을 업데이트하는 방법을 설명

▼ 그림 3 | 은닉층의 오차 업데이트



6 은닉층의 오차 수정하기

- **은닉층의 오차 수정하기**
  - 마찬가지로 가중치에 기울기를 뺀 값을 구해야 함
  - 이때 우리가 구하려는 값이 w11이므로 다음과 같이 계산

$$w_{11}(t+1) = w_{11}t - \frac{\partial 오차 Y_{\text{out}}}{\partial w_{11}}$$

6 은닉층의 오차 수정하기

- **은닉층의 오차 수정하기**
  - 여기서 $\frac{\partial 오차 Y_h}{\partial w_{11}}$ 가 아니라 $\frac{\partial 오차 Y_{out}}{\partial w_{11}}$ 이라는 점을 주목하기 바람
  - 그 이유는 Y_h가 은닉층 안에 위치해 있으므로 겉으로 드러나지 않기 때문임
  - 그 값을 알 수 없음
  - 우리가 알 수 있는 출력 값은 Y_{out}뿐이므로 은닉층의 오차 업데이트를 위한 기울기를 구할 때도 Y_{out}에서 출발해야 함

6 은닉층의 오차 수정하기

- **은닉층의 오차 수정하기**
  - 이제 앞서 계산했던 바와 마찬가지로 기울기에 해당하는 $\frac{\partial 오차 Y_{out}}{\partial w_{11}}$ 을 구해 보자
  - 체인 룰을 적용해 다음과 같이 계산

$$\frac{\partial 오차 Y_{out}}{\partial w_{11}} = \underbrace{\frac{\partial 오차 Y_{out}}{\partial y_{h1}}}_{①} \cdot \underbrace{\frac{\partial y_{h1}}{\partial 가중합_1}}_{②} \cdot \underbrace{\frac{\partial 가중합_1}{\partial w_{11}}}_{③}$$

6 은닉층의 오차 수정하기

- 은닉층의 오차 수정하기
  - 여기서 ❷ 항과 ❸항은 이전과 같은 방법으로 계산
  - 다음과 같이 바꾸어 줄 수 있음

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

7 은닉층의 오차 계산법

7 은닉층의 오차 계산법

- **은닉층의 오차 계산법**
  - 은닉층에서 ❶항은 계산이 조금 다름
  - 오차 Y_{out}안에는 오차 y_{o1}과 오차 y_{o2}가 포함되어 있음
  - 이전에는 오차 Y_{out}을 y_{o1}에 의해 편미분할 때 y_{o1}과 관계없는 항인 오차 y_{o2}는 상수가 되어 사라졌음
  - 남는 것은 $\frac{\partial 오차\, y_{o1}}{\partial y_{o1}}$ 뿐이었음

7 은닉층의 오차 계산법

- 은닉층의 오차 계산법
  - 이번에는 y_{h1}에 대해 미분해야 함
  - Y_{h1}은 오차 y_{o1}과 오차 y_{o2}의 형성에 모두 관계가 있음
  - $\frac{\partial 오차 y_{o1}}{\partial y_{h1}}$ 과 $\frac{\partial 오차 y_{o2}}{\partial y_{h1}}$ 가 모두 계산되어야 하므로 계산이 다음과 같이 조금 복잡해짐

$$\textcircled{1}\quad \frac{\partial 오차 Y_{out}}{\partial h_1} = \frac{\partial(오차 y_{o1} + 오차 y_{o2})}{\partial y_{h1}} = \underbrace{\frac{\partial 오차 y_{o1}}{\partial y_{h1}}}_{ⓐ} + \underbrace{\frac{\partial 오차 y_{o2}}{\partial y_{h1}}}_{ⓑ}$$

7 은닉층의 오차 계산법

- 은닉층의 오차 계산법
  - 먼저 a 부분을 보자
  - 체인 룰에 의해 다음과 같이 바뀜

$$\frac{\partial 오차\, y_{o1}}{\partial y_{h1}} = \underbrace{\frac{\partial 오차\, y_{o1}}{\partial 가중합_3}}_{a\text{-}1} \cdot \underbrace{\frac{\partial 가중합_3}{\partial y_{h1}}}_{a\text{-}2}$$

7 은닉층의 오차 계산법

- **은닉층의 오차 계산법**
  - 이 중 a_{-1} 부분을 다시 미분하면 역시 체인 룰에 의해 다음과 같이 바뀜

$$\frac{\partial 오차 y_{o1}}{\partial 가중합_3} = \frac{\partial 오차 y_{o1}}{\partial y_{o1}} \cdot \frac{\partial y_{o1}}{\partial 가중합_3}$$

7 은닉층의 오차 계산법

- **은닉층의 오차 계산법**
  - 이 중 먼저 $\frac{\partial 오차 y_{o1}}{\partial y_{o1}}$ 에 포함된 오차 y_{o1}은 $\frac{1}{2}(y_{t1}-y_{o1})^2$ 이므로 $\frac{\partial 오차 y_{o1}}{\partial y_{o1}}$ 을 y_{o1}로 편미분하면 $y_{o1} - y_{t1}$이 됨
  - $\frac{\partial y_{o1}}{\partial 가중합_3}$ 은 앞서 설명한 대로 시그모이드 함수의 미분
  - $y_{o1} \cdot (1 - y_{o1})$로 계산
  - 이제 나머지 a $_{-2}$를 미분하면 w_{31}이 남음
  - a $_{-1}$, a $_{-2}$ 를 정리하면 a 는 다음과 같음

$$\frac{\partial 오차 y_{o1}}{\partial y_{h1}} = (y_{o1} - y_{t1}) \cdot y_{o1}(1 - y_{o1}) \cdot w_{31}$$

7 은닉층의 오차 계산법

- 은닉층의 오차 계산법
  - 여기서 $(y_{o1} - y_{t1}) \cdot y_{o1}(1-y_{o1})$ 부분이 눈에 익지 않은가?
  - 앞서 기억해 두었던 델타 식(δy)의 형식
  - 지금 우리는 y_{o1}을 구해야 하므로 델타식을 δy_{o1}이라고 할 때, 앞의 값은 다음과 같이 간단하게 표시할 수도 있음

$$\frac{\partial 오차 y_{o1}}{\partial y_{h1}} = \delta y_{o1} \cdot w_{31} \cdots a'$$

  - 이제 b 부분을 볼까?
  - 역시 체인 룰에 의해 다음과 같이 변형

$$\frac{\partial 오차 y_{o2}}{\partial y_{h1}} = \frac{\partial 오차 y_{o2}}{\partial 가중합_4} \cdot w_{41}$$

7 은닉층의 오차 계산법

- 은닉층의 오차 계산법
  - 이 중 $\frac{\partial 오차 y_{o2}}{\partial 가중합_4}$ 부분은 체인 룰에 의해 $\frac{\partial 오차 y_{o2}}{\partial y_{o2}} \cdot \frac{\partial y_{o2}}{\partial 가중합_4}$ 로 바뀜
  - 앞서 식 ⓐ 풀이에서 설명한 방식과 똑같이 적용되므로 답은 바로 나옴

$$\frac{\partial 오차 y_{o2}}{\partial y_{o2}} = (y_{o2} - y_{t2}) \cdot y_{o2}(1 - y_{o2}) \cdot w_{41}$$

  - 여기서 $(y_{o2} - y_{t2}) \cdot y_{o2}(1 - y_{o2})$ 부분은 델타식 형식
  - 이 델타식을 δy_{o2} 라고 할 때 주어진 식은 다음과 같이 표시할 수 있음

$$\frac{\partial 오차 y_{o2}}{\partial y_{h1}} = \delta y_{o2} \cdot w_{41} \quad \cdots \text{ⓑ}'$$

7 은닉층의 오차 계산법

- **은닉층의 오차 계산법**
  - 이제 ⓐ′와 ⓑ′를 이용해 ❶을 다시 정리하면 다음과 같음

$$\frac{\partial 오차 Y_{out}}{\partial h_1} = \frac{\partial 오차\, y_{o1}}{\partial y_{h1}} + \frac{\partial 오차\, y_{o2}}{\partial y_{h1}} = \delta y_{o1} \cdot w_{31} + \delta y_{o2} \cdot w_{41}$$

  - 이제 이 값을 이용해 6절에 나온 은닉층의 오차 업데이트 식을 완성하면 다음과 같음

$$\begin{aligned}\frac{\partial 오차 Y_{out}}{\partial w_{11}} &= \frac{\partial 오차 Y_{out}}{\partial y_{h1}} \cdot \frac{\partial y_{h1}}{\partial 가중합_1} \cdot \frac{\partial 가중합_1}{\partial w_{11}} \\ &= (\delta y_{o1} \cdot w_{31} + \delta y_{o2} \cdot w_{41}) y_{h1} (1 - y_{h1}) \cdot x_1\end{aligned}$$

8 델타식

8 델타식

● 델타식

- 이제 출력층과 은닉층의 업데이트를 위해 도출된 두 개의 식을 비교해 보자

  - 출력층의 오차 업데이트 $\boxed{(y_{o1} - y_{t1}) \cdot y_{o1}(1 - y_{o1})} \cdot y_{h1}$
  - 은닉층의 오차 업데이트 $\boxed{(\delta y_{o1} \cdot w_{31} + \delta y_{o2} \cdot w_{41}) y_{h1}(1 - y_{h1})} \cdot x_1$

8 델타식

- 델타식
  - 여기서 박스로 표시한 두 부분을 비교해 보자
  - (y_{o1}-y_{t1})이 (δy_{o1} · w_{31} +δy_{o2} · w_{41})로 바뀌었지만, 나머지 부분은 out(1 - out) 형태를 띠고 있음
  - 여기서 (y_{o1}- y_{t1})은 오차 값
  - 은닉층에서는 이렇게 오차를 계산할 수 없음
  - 은닉층에서 일어나는 일은 우리 눈으로 볼 때는 알 수 없으므로 출력층에서 y_o값을 가져와서 계산해야 함
  - 앞 식의 δy_{o1} · w_{31} + δy_{o2} · w_{41})처럼 형태가 복잡해졌을 뿐 결국 오차를 나타냄
  - 두 식 모두 ‘오차 · out(1 - out)’ 형태, 즉 델타식의 형태로 단순화할 수 있음

8 델타식

- 델타식
  - 델타식이 중요한 이유는 이렇게 한 층을 거슬러 올라갈 때마다 같은 형태가 계속 나타나기 때문임
  - 델타식을 파악하고 나면 이를 코딩으로 설계하는 것도 어렵지 않음
  - 은닉층의 델타식이므로 이것을 δh 라고 할 때, 은닉층의 가중치 업데이트를 식으로 표현하면 다음과 같음

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

  - 이렇게 해서 모든 출력층과 은닉층의 가중치가 각각 업데이트되는 과정을 수식을 통해 살펴보았음