

## 4.4.2 신경망에서의 기울기

여기서 '기울기'란 가중치 매개변수에 대한 손실 함수의 기울기.

$\frac{\partial L}{\partial w}$ 의 각 원소는 '각각의 원소에 대한 편미분'이다.

이때,  $\frac{\partial L}{\partial w}$ 의 행상이  $w$ 와 같음.

→  $\frac{\partial L}{\partial w}$ 의  $\frac{\partial L}{\partial w_{11}}$ 은 대략 0.2임.

이는  $w_{11}$ 을 1만큼 늘리면 손실함수값은 0.2만큼 증가한다는 의미

→  $\frac{\partial L}{\partial w}$ 의  $\frac{\partial L}{\partial w_{23}}$ 은 대략 -0.5임

이는  $w_{23}$ 을 1만큼 늘리면 손실함수값은 0.5만큼 감소한다는 의미

그래서 '손실함수를 줄인다'는 관점에서는

1.  $\frac{\partial L}{\partial w_{11}}$ 은 양의 방향으로, 2.  $\frac{\partial L}{\partial w_{23}}$ 은 음의 방향으로 갱신 해야함.

또, 한번에 갱신되는 양이  $\frac{\partial L}{\partial w_{23}}$ 이  $w_{11}$ 보다 '크게'기어 한다는 것 또한 수 있음.

## 4.4.5 (학습 알고리즘 구현하기)

### 전제

신경망에는 적응 가능한 '가중치' '편향'이 있음.

이를 훈련 데이터에 적응하도록 조정하는 과정을 '학습'이라 함.

### 1 - 미니배치

'훈련데이터' 중 일부를 무작위로 가져와 <sup>오늘</sup> '미니배치'함.

그 미니배치의 '손실함수 값'을 줄이는데 목표.

### 2 - 기울기 산출

각 '가중치 매개변수'의 기울기를 구함.

'기울기'는 손실함수의 값을 가장 작게 하는 방향으로 제시하기 때문

### 3 - 매개변수 갱신

'가중치 매개변수'를 기울기 방향으로 '아주 조금' 갱신함

### 4 - 반복

#### 1-3 반복

이는 '경사 하강법'으로 매개변수를 갱신하는 방법임.

→ 이를 '확률적 경사 하강법' (SGD)라 함.