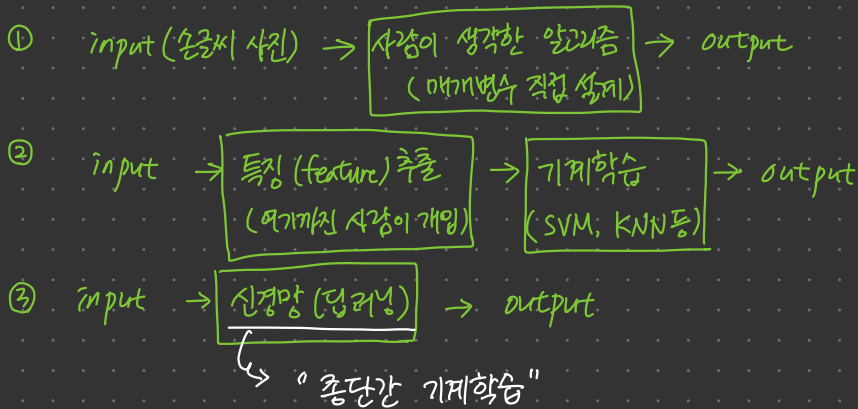


# Chapter 4. 신경망 학습

## 1. 데이터에서 학습한다!

- 기계학습  $\Rightarrow$  사람의 개입을 최소화하고, 수집한 데이터로부터 패턴을 찾으려 시도.  
 $\Rightarrow$  "데이터 주도 학습"

ex) 손글씨 숫자 5를 찾는 알고리즘



: 입력부터 출력까지를 사람의 개입 없이 연산.

- 훈련 데이터 - 시험 데이터

70%		30%
-----	--	-----

$\Rightarrow$  모델의 "범용 능력" 판단. (기계학습의 최종 목표임!)

아직 보지 못한 데이터 (훈련 X)로도  
문제를 올바르게 풀어내는 능력.

\* 오버 피팅 (overfitting) : 한 데이터셋에만 지나치게 최적화된 상태.

## 2. 손실 함수

→ 신경망에서 최적의 매개변수 탐색을 위한 하나의 지표

• 평균 제곱 오차 (MSE)

$$E = \frac{1}{2} \sum_k (y_k - t_k)^2$$

( $y_k$ : 출력값(추정값),  $t_k$ : 정답 레이블,  $k$ : 데이터 차원의 수)

⇒ 추정값과 참값의 차이 제곱합.

\* 원-핫 인코딩: 한 원소만 1로 하고, 그 외는 0으로 나타내는 표기법.

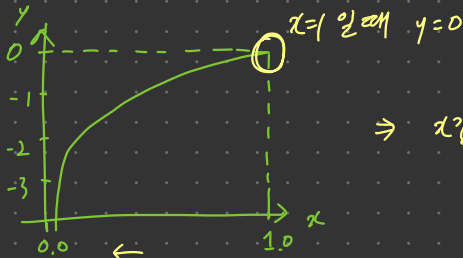
• 교차 엔트로피 오차 (CEE)

$$E = - \sum_k t_k \log y_k$$

(log: 밑이 2인 자연로그)

⇒ 정답일 때의 추정 ( $t_k=1$  일 때의  $y_k$ )의 자연로그

$y = \log x$  그래프



⇒ x값이 작아질수록 (정답일 때의 출력값이 작아질수록)

$|y|$  (오차값)이 커진다.

x가 0에 가까워질수록 y값 감소.

• 미니배치 학습에서의 손실함수

$$E = - \frac{1}{N} \sum_n \sum_k t_{nk} \log y_{nk}$$

⇒ n개의 데이터를 확장

데이터의 수가 아주 많은 경우,

일부 데이터만 추출 전체의 '공식치'로 사용.

• 왜 손실함수를 설정하는가?

↳ 연속적으로 변화하는 값.

매개변수 미분 갱신 때문임.

정확도는 불연속적인 값으로 조정되어야 미분하면 0이 될 수 있다.

ex) 33% → 34%.

### 3. 수치 미분

\* 복습) 미분 : '특정 순간'의 변화량  $\frac{df(x)}{dx} = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$

→ 파이썬에 적용하려면 '반올림 오차' 발생

:  $1e-50$  을 float32형으로 나타내면 0.0 이 된다.  
(값이 너무 작아서 문제)

$h \rightarrow 0$  으로 무한히 줄일 수 없으므로 '정확한 미분'이 될 수 없다.

⇒ sol) "수치미분" : 아주 작은 채움으로 미분  $\lim_{h \rightarrow 0} \frac{f(x+h) - f(x-h)}{2h}$

• 편미분 : 변수가 여러 개인 함수에 대한 미분 ⇒ 목표 변수 하에 초점을 맞추고, 다른 변수는 값을 고정한다.

ex)  $f(x_0, x_1) = x_0^2 + x_1^2$

$$\begin{cases} \frac{\partial f}{\partial x_0} = 2x_0 + x_1^2 \\ \frac{\partial f}{\partial x_1} = x_0^2 + 2x_1 \end{cases}$$

### 4. 기울기

• 기울기 (gradient) : 모든 변수의 편미분을 벡터로 정리한 것.

ex)  $(\frac{\partial f}{\partial x_0}, \frac{\partial f}{\partial x_1})$  :  $x_0, x_1$  의 편미분을 동시에 계산

⇒ 기울기는 벡터이므로 **방향**을 가짐. ⇒ 손실함수의 최솟값을 가리킴.

• 경사법 (경사하강법) : 기울기 이용하여 함수의 값을 점차 줄여나가는 것.

(1회갱신)

$$\begin{cases} x_0 = x_0 - \eta \frac{\partial f}{\partial x_0} \\ x_1 = x_1 - \eta \frac{\partial f}{\partial x_1} \end{cases}$$

$\eta$  (eta, 에타) : 학습률, learning rate, 갱신하는 양

→ 하이퍼 파라미터 (사람이 직접 설정)

⇒ 갱신 반복 (step-num)

· 신경망에서의 기울기 (= 가중치 매개변수에 대한 손실함수의 기울기)

ex) 가중치  $W = \begin{pmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{pmatrix}$ , 손실함수  $L$

$\Rightarrow \frac{\partial L}{\partial W} = \begin{pmatrix} \frac{\partial L}{\partial w_{11}} & \frac{\partial L}{\partial w_{12}} & \frac{\partial L}{\partial w_{13}} \\ \frac{\partial L}{\partial w_{21}} & \frac{\partial L}{\partial w_{22}} & \frac{\partial L}{\partial w_{23}} \end{pmatrix}$   $\Rightarrow$  손실함수를 줄이기 위해 기여하는 값.

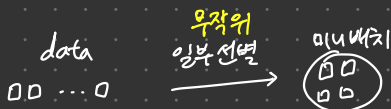
$\Rightarrow$  각 원소의 편미분 (변화량)

## 5. 학습 알고리즘 구현하기

<정리> 신경망 학습 절차  $\Rightarrow$  "확률적 경사 하강법 (SGD)"

0) 신경망에는 '가중치'와 '편향'이 있다. 이를 조정하면서 혼돈데이터에 적응시킴 ('학습')

1) 미니배치



목적: 미니배치의 손실값을 줄이는 것.

2) 기울기 산출

각 가중치 매개변수 ( $W$ )의 기울기  $\Rightarrow$  손실함수의 값을 가장 작게 하는 방향 제시

3) 매개변수 갱신

$W$ 를 기울기 방향으로 "아주 조금 갱신"

4) 1~3 반복