

Machine Learning

Chapter 4

신경망 학습

4. 신경망 학습

4.1 데이터에서 학습한다

4.2 손실 함수

4.3 수치 미분

4.4 기울기

4.5 학습 알고리즘 구현

4.6 정리

4.1 데이터에서 학습한다

■ 사람

- 경험과 직관을 단서로 일을 진행

■ 기계학습

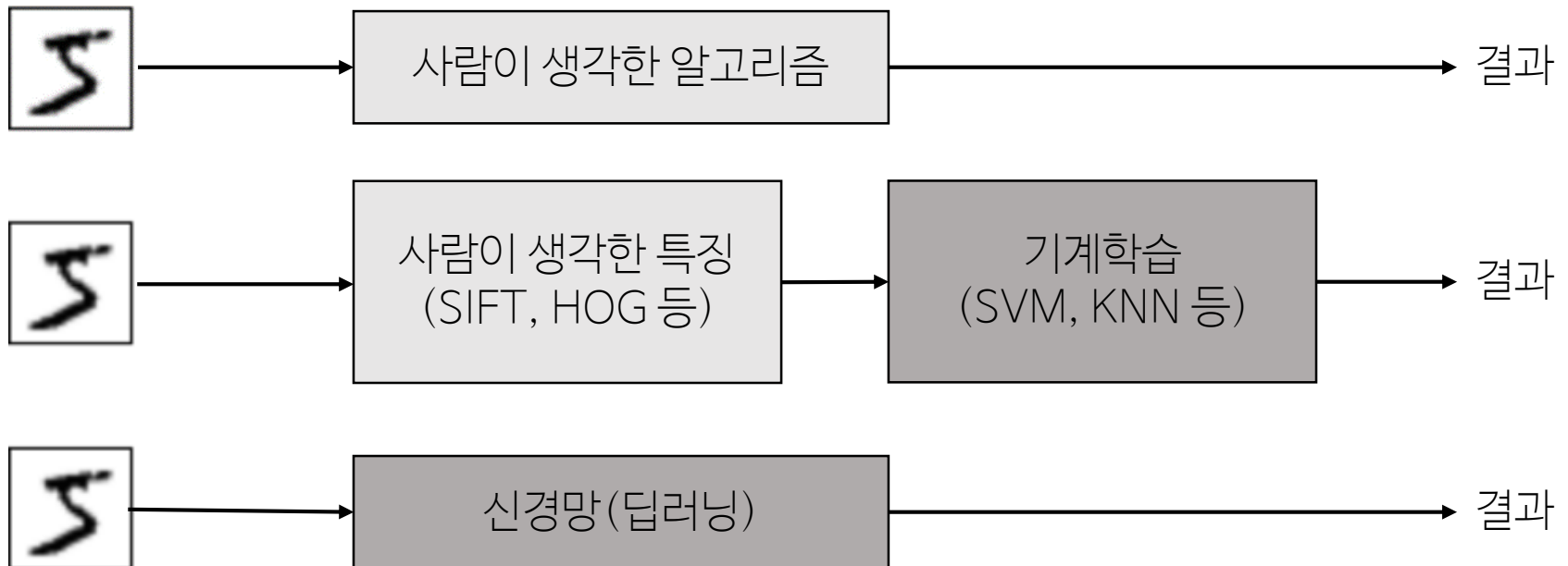
- 사람의 개입을 최소화
- 수집한 데이터로부터 패턴을 찾음.

선형 분리 가능 문제는 유한번의 학습을 통해 풀 수 있다는 사실이 퍼셉트론 수렴 정리
perceptron convergence theorem 로 증명되었습니다.

하지만 비선형 분리 문제는 자동으로 학습할 수 없습니다.

4.1 데이터에서 학습한다

■ 데이터 주도 학습



딥러닝을 종단간 기계학습이라고도 한다. 여기서 종단간은 처음부터 끝까지라는 의미로, 데이터(입력)에서 목표한 결과를 사람의 개입 없이 얻는다는 뜻을 담고 있다.

4.1 데이터에서 학습한다

■ **훈련 데이터** training data 와 **시험 데이터** test data

- 훈련 데이터만 사용하여 학습하면서 최적의 매개변수를 찾은 후에 시험 데이터를 사용하여 모델을 평가

■ **훈련데이터와 시험데이터를 나누는 이유**

- 훈련 데이터에 포함되지 않은 데이터로도 문제를 올바르게 해결할 수 있는 능력(**범용 능력**)이 있는지 확인하기 위해서
- 데이터셋 하나로만 매개변수의 학습과 평가를 수행하면 올바른 평가가 될 수 없음. 한 데이터셋에만 지나치게 최적화된 상태를 **오버피팅** overfitting 이라고 함.

4.2 손실 함수

■ 손실 함수 loss function

- 신경망에서 최적의 매개변수를 탐색하는데 사용되는 하나의 지표

■ 손실 함수의 종류

- 평균 제곱 오차
- 교차 엔트로피 오차

손실 함수는 신경망 성능의 '나쁨'을 나타내는 지표로, 현재의 신경망이 훈련 데이터를 얼마나 잘 처리하지 못하는지를 나타낸다.

'나쁨을 최소화 하는것'이 '좋음을 최대화 하는것'과 같기 때문에 본질적으로 수행하는 일은 다르지 않음.

4.2 손실 함수

■ 평균 제곱 오차 mean squared error, MSE

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

- 예) 숫자 0~9 까지의 신경망의 출력
 - $y = [0.1, 0.05, 0.6, 0.0, 0.05, 0.1, 0.0, 0.1, 0.0, 0.0]$
 - $t = [0, 0, 1, 0, 0, 0, 0, 0, 0, 0] \rightarrow$ '원-핫 인코딩'
 - y 는 소프트맥스 함수의 출력, t 는 정답레이블

■ 교차 엔트로피 오차 cross entropy error, CEE

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

4.2 손실 함수

■ 미니배치mini-batch 학습

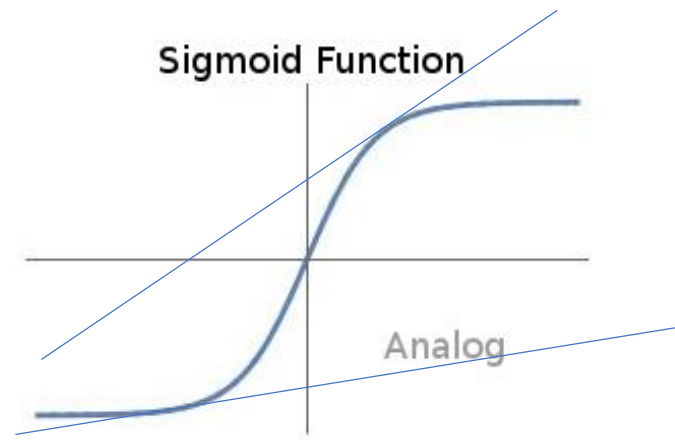
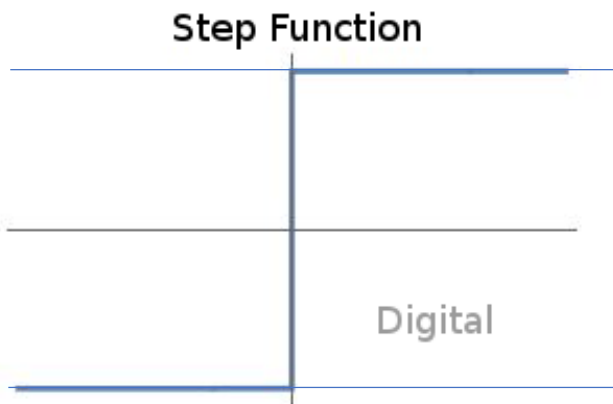
- $$E = -\frac{1}{N} \sum_n \sum_k t_{nk} \log y_{nk}$$

- 모든 훈련데이터를 대상으로 손실함수의 합을 구하려면 시간이 많이 걸림
- 전체의 데이터 중에서 무작위로 일부를 뽑아 그 일부만으로 학습하는 방법
- 일부 표본 데이터로 전체를 비슷하게 예측하는 방법

4.2 손실 함수

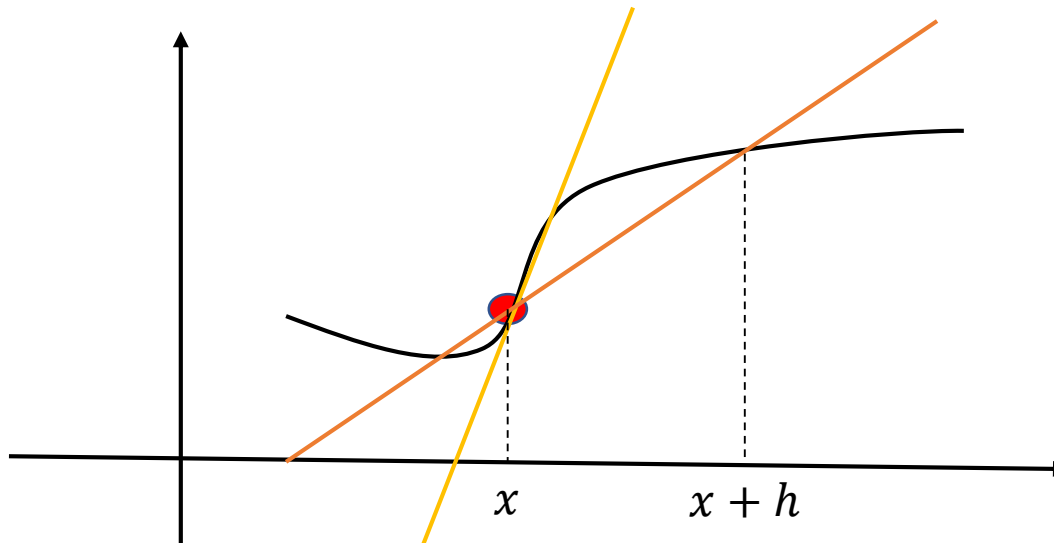
■ 손실 함수를 사용하는 이유?

- 미분을 통하여 매개변수(가중치와 편향)를 움직여 손실 함수의 값을 변화시키는데 정확도를 지표로 삼게 되면 미분 값이 대부분의 장소에서 0이 되어 매개변수를 갱신할 수 없기 때문이다.



4.3 수치 미분

- $$\frac{df(x)}{dx} = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$



이처럼 아주 작은 차분으로 미분하는 것을 수치 미분이라고 함. 수식을 전개해 미분하는 것은 해석적이라는 말을 이용하여 해석적 해 등으로 표현된다. 해석적 미분은 오차를 포함하지 않은 진정한 미분값을 구해준다.

4.4 기울기

■ 기울기란?

- 모든 변수의 편미분을 벡터로 정리한 것
- 기울기를 벡터로 그려내면 기울기는 함수의 가장 낮은 장소 (최솟값)를 가리키게 된다.
- 즉, 기울기가 가리키는 쪽은 각 장소에서 함수의 출력 값을 가장 크게 줄이는 방법

4.4 기울기

■ 경사법

- 현 위치에서 기울어진 방향으로 이동 후 기울기를 구하고 다시 나아가기를 반복하여 함수의 값을 점차 줄이는 방법

- $x = x - \eta \frac{\partial f}{\partial x}$

함수가 극솟값 최솟값 그리고 **안장점**이 되는 지점에서 기울기는 0이 된다.

극솟값은 국소적인 최솟값인 점이고 안장점은 어느 방향에서 보면 극댓값이고 다른 방향에서 보면 극솟값이 되는 점이다.

기울기가 0인 장소를 찾지만 그것이 반드시 최솟값이라고 할 수 없다.

복잡하고 찌그러진 모양의 함수라면 평평한 곳으로 파고들면서 **고원**이라고 하는 학습이 진행되지 않는 정체기에 빠질 수 있다.

4.5 학습 알고리즘 구현하기

- 전제

- 신경망에는 적응 가능한 가중치와 편향이 있고, 이 가중치와 편향을 훈련 데이터에 적응하도록 조정하는 과정을 학습이라 한다. 신경망 학습은 다음과 같이 4단계로 수행한다.

- 1단계-미니배치

- 훈련 데이터 중 일부를 무작위로 가져온다. 이렇게 선별한 데이터를 미니배치라 하며, 그 미니배치의 손실 함수 값을 줄이는 것이 목표이다.

- 2단계-기울기 산출

- 미니배치의 손실 함수 값을 줄이기 위해 각 가중치 매개변수의 기울기를 구한다. 기울기는 손실 함수의 값을 가장 작게 하는 방향을 제시한다.

- 3단계-매개변수 갱신

- 가중치 매개변수를 기울기 방향으로 아주 조금 갱신한다.

- 4단계-반복

- 1~3단계를 반복한다.

4.6 정리

이번 장에서 배운 내용

- 기계학습에서 사용하는 데이터셋은 훈련 데이터와 시험 데이터로 나눠 사용한다.
- 훈련 데이터로 학습한 모델의 범용 능력을 시험 데이터로 평가한다.
- 신경망 학습은 손실 함수를 지표로, 손실 함수의 값이 작아지는 방향으로 가중치 매개변수를 갱신한다.
- 가중치 매개변수를 갱신할 때는 가중치 매개변수의 기울기를 이용하고, 기울어진 방향으로 가중치의 값을 갱신하는 작업을 반복한다.
- 아주 작은 값을 주었을 때의 차분으로 미분하는 것을 수치 미분이라고 한다.
- 수치 미분을 이용해 가중치 매개변수의 기울기를 구할 수 있다.
- 수치 미분을 이용한 계산에는 시간이 걸리지만, 그 구현은 간단하다. 한편, 다음 장에서 구현하는 오차역전파법은 기울기를 고속으로 구할 수 있다.