

Chapter 08. AI의 학습원리와 데이터 처리

- AI 모델에서 학습은 어떻게 이루어지나?



8-1. 머신러닝 모형의 원리 : 분류 모형

(출처: K-ICT 딥러닝 개요, 송준이(아이덴티파이) 자료 인용)

■ 센서 데이터의 분류

- 2개의 입력 센서 데이터로 고장 유무 판별

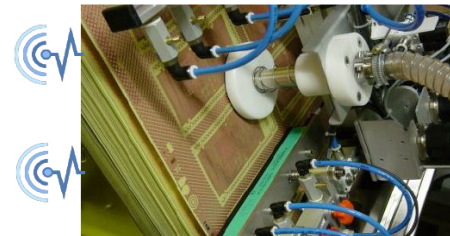
x_1 : 첫 번째 센서에서 수집된 값

x_2 : 두 번째 센서에서 수집된 값

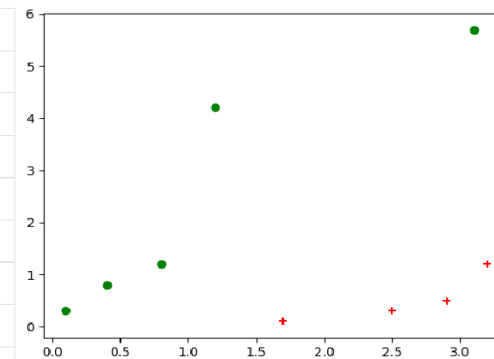
Y : 장치의 오작동 여부

x_1

x_2



x_1	x_2	y
0.1	0.3	False
0.8	1.2	False
2.5	0.3	True
1.2	4.2	False
0.4	0.8	False
3.2	1.2	True
1.7	0.1	True
3.1	5.7	False
2.9	0.5	True



<그림> 2개의 센서 데이터와 장비 고장 유무

8-1. 머신러닝 모형의 원리 : 분류 모형

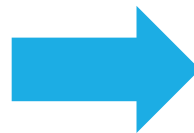
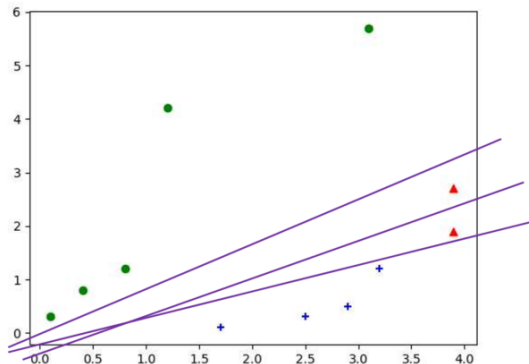
(출처: K-ICT 딥러닝 개요, 송준이(아이덴티파이) 자료 인용)

■ 센서 데이터의 분류

- 새로운 센서 데이터가 수집된 경우, 장치의 오작동 여부 예측

x1	x2	y
3,9	2,7	?
3,9	1,9	?

분류 문제 :
2개의 데이터를 가장 잘 구별하는 선형
판별식을 찾는 과정



$$w_1x_1 + w_2x_2 = b$$

8-1. 머신러닝 모형의 원리 : 분류 모형

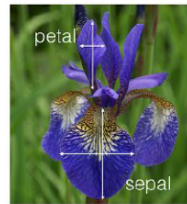
■ IRIS 꽃 분류

- 입력변수

- 꽃받침 길이(sepal length)
- 꽃받침 너비(sepal width)
- 꽃잎 길이(petal length)
- 꽃잎 너비(petal width)

- 출력변수

- IRIS의 세 품종
- 세토사(Setosa)
- 버시컬러(Versicolour)
- 버지니카(Virginica)



특성(feature)

레이블(label)

x1	x2	x3	x3	y
sepal length	sepal width	petal length	petal width	species
5.0	3.4	1.6	0.4	Iris-setosa
5.6	2.9	3.6	1.3	Iris-versicolor
6.3	2.8	5.1	1.5	Iris-virginica
...
...

표. 학습 데이터 예시

- 학습이란

학습 데이터에 대해 아래를 만족하는 최적의 파라미터 $\{(w_1, w_2, w_3) b\}$ 를 찾는 과정

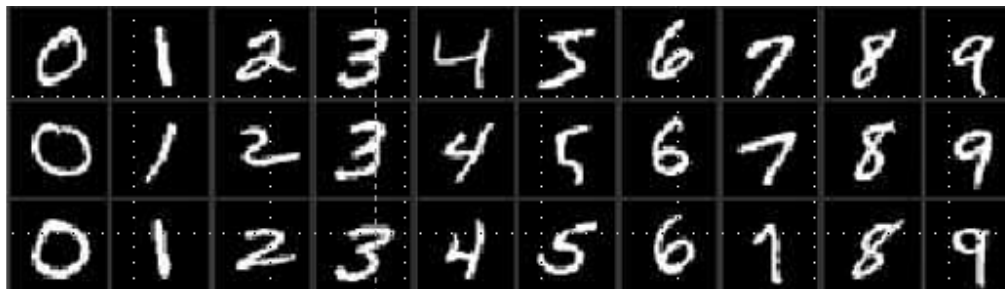
$$f(w_1x_1 + w_2x_2 + w_3x_3 + b) \Rightarrow 0 \text{ or } 1 \text{ or } 2$$

8-2. 머신러닝 모형의 원리 : 분류 모형(이미지 데이터 세트)

(출처: K-ICT 딥러닝 개요, 송준이(아이덴티파이) 자료 인용)

■ MNIST 필기체 분류

- 문제
 - 사람이 필기체로 작성한 0부터 9까지의 숫자 이미지에 대해 적힌 숫자 값을 식별하기
- 학습 데이터
 - 입력값 : 0부터 9까지 필기체로 작성된 숫자 이미지
 - 레이블 : 이미지에 적힌 실제 숫자 값



<그림> MNIST 필기체 이미지

학습정리 #8차

■ 인공지능 모델에서의 학습

- 훈련(Training) 데이터 세트를 이용해서 학습을 진행함
- “명시적인 조건”이 없이, 라벨링된 데이터를 이용해서 스스로 판단 기준을 만들

■ 분류 모형(Classification Model)

- 대상의 타입(Type)을 스스로 찾아내고 이를 근거로 각 대상을 적절하게 소팅(Sorting)
(비고 : 대상의 값(Value)의 예측 - 회귀모형)
- 군집모형 분석과 같이 데이터 라벨링이 없더라도 판정이 가능함

■ 분류모형의 사용 예

- 데이터의 포맷과 상관 없이 다양한 데이터 소스에 맞도록 모델 개발 가능
- 스팸메일 필터(텍스트), 스마트 팩토리의 고장 제품 판단(이미지), 음성인식 비서(시리, S-Voice) 등