

10

강

인공지능

머신러닝(1)

컴퓨터과학과 이병래교수

학습목차

1 학습이란 무엇인가?

2 귀납적 학습

3 결정트리 학습





학습이란 무엇인가?

1. 머신러닝의 개요

■ 머신러닝(machine learning)

- 인공지능 시스템이 지능적 행동능력을 갖기 위해서 외부 환경으로부터의 정보를 이용하여 시스템 내부에 지식을 형성하고 저장하는 과정

“시스템이 동일한 대상에 대한 작업을 다음 번에는 더욱 효율적이고 더욱 효과적으로 수행하도록 하는 시스템의 적응적 변화”

– Herbert Simon

2. 학습의 유형

■ 지도학습(supervised learning)

- 입력과 이 입력에 대해 기대하는 출력(**레이블**)을 학습 데이터로 제시함
- 입력에 대한 시스템의 출력이 기대하는 출력과 같아지도록 시스템을 변화시킴

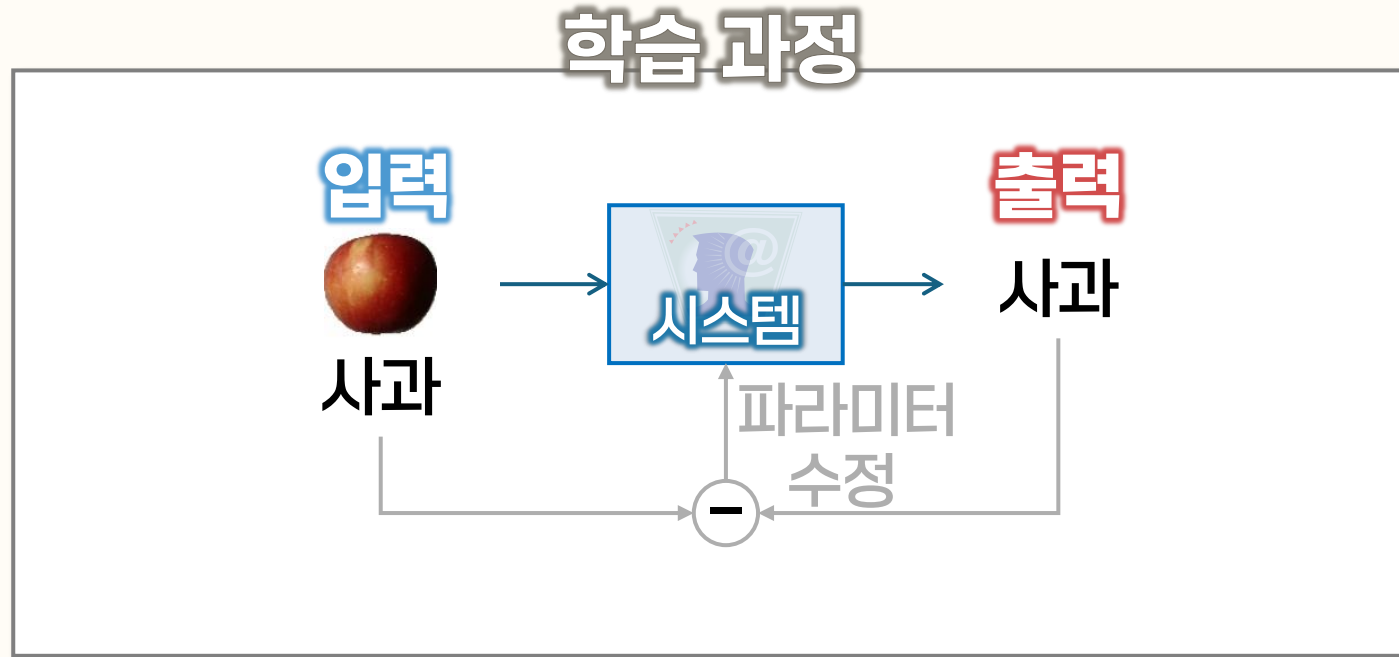
학습 표본 집합



2. 학습의 유형

■ 지도학습(supervised learning)

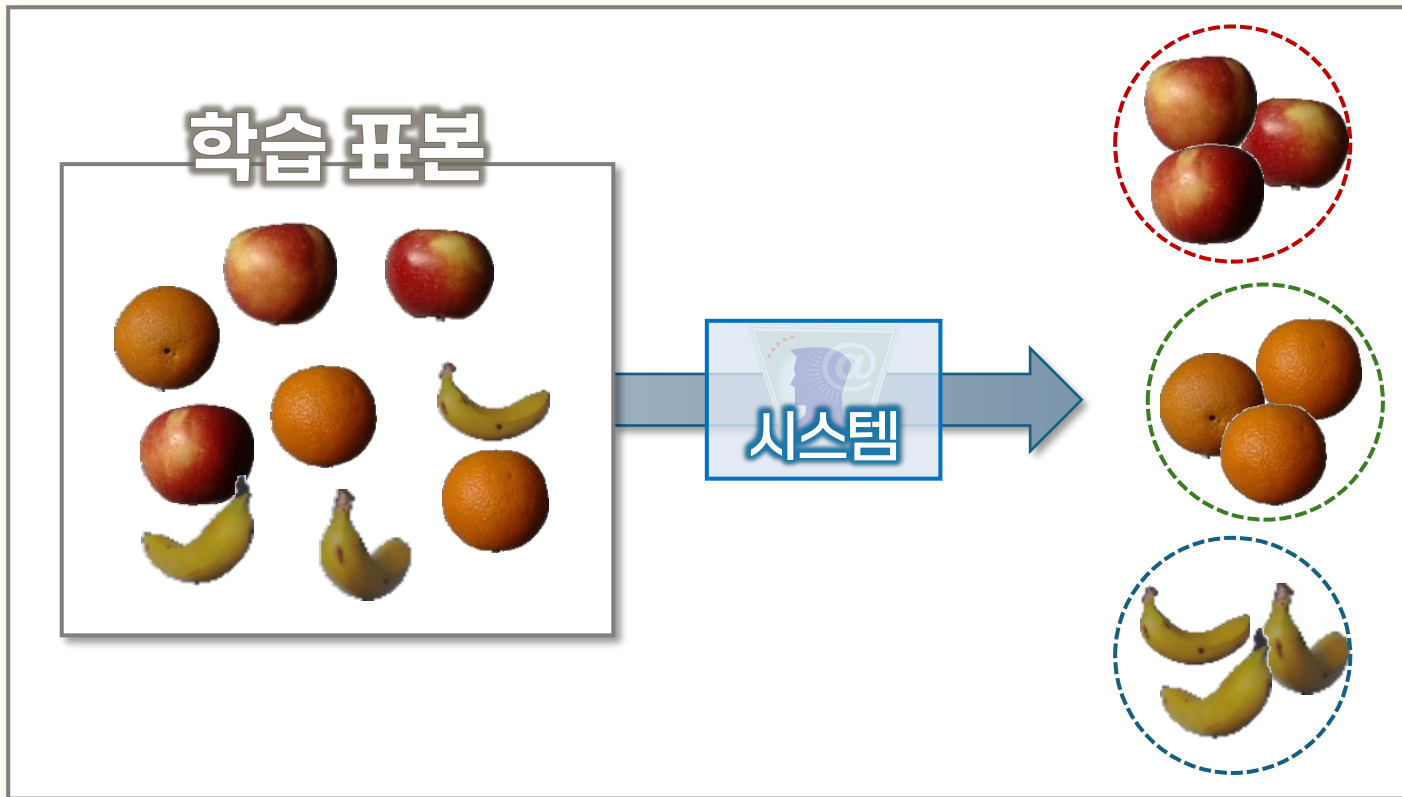
- 입력과 이 입력에 대해 기대하는 출력(**레이블**)을 학습 데이터로 제시함
- 입력에 대한 시스템의 출력이 기대하는 출력과 같아지도록 시스템을 변화시킴



2. 학습의 유형

■ 비지도학습(unsupervised learning)

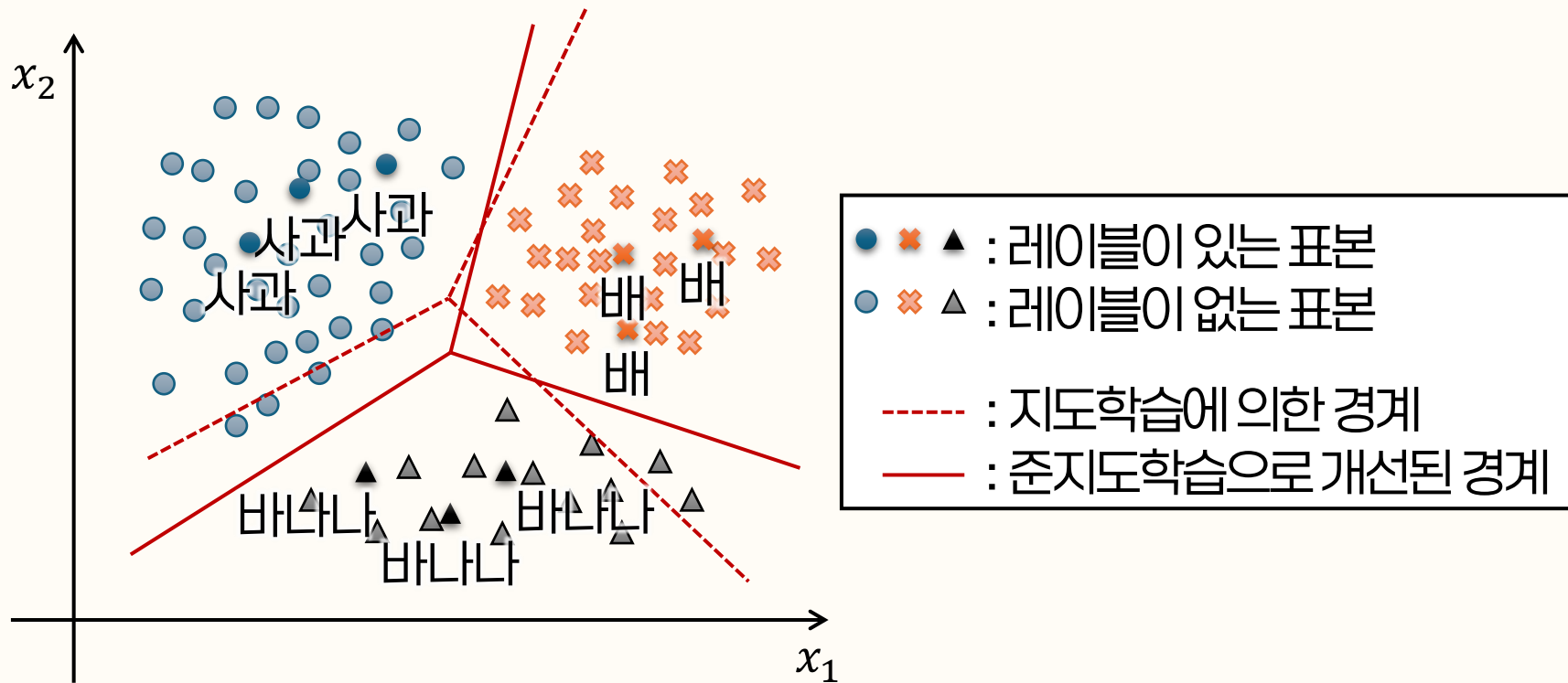
- 학습 데이터는 시스템에 대한 입력 값만으로 구성
- 유사한 입력에 대해서는 동일한 출력을 내도록 학습함



2. 학습의 유형

준지도학습(semi-supervised learning)

- 대량의 레이블이 없는 데이터와 상대적으로 적은 수의 레이블이 지정된 데이터를 학습에 사용



2. 학습의 유형

❑ 자기지도학습(self-supervised learning)

- 레이블이 없는 데이터만으로 학습
- 학습 알고리즘은 이 데이터 집합에 대해 스스로 레이블을 자동 생성

예 문장에서 일부 단어를 가린 다음 원래의 문장을 레이블로 사용

➡ 가려진 단어를 예측하는 학습을 할 수 있음

2. 학습의 유형

■ 전이학습(transfer learning)

- 어떠한 문제를 해결하기 위해 학습된 결과를 이와 특성이 유사한 다른 응용 학습에 활용하는 학습 방법
- 목표 시스템에 필요한 요소를 추가한 후 기존 학습 결과를 미세조정함

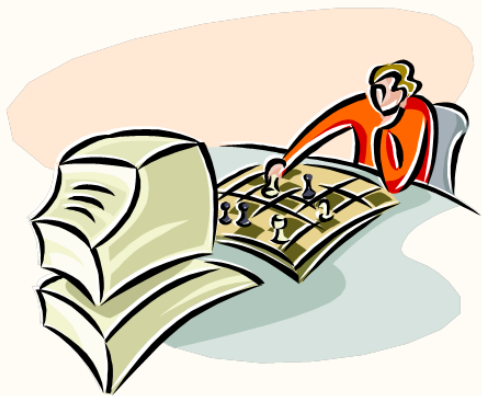
예 **YOLO(2015)**: 영상에서 객체 위치와 클래스를 빠르게 계산할 수 있는 실시간 객체 검출 시스템

- ImageNet 데이터셋을 이용하여 사전학습
 - ➡ 사전학습 모델에 객체 검출을 위한 요소를 추가한 후 PASCAL VOC 데이터셋으로 학습

2. 학습의 유형

■ 강화학습(reinforcement learning)

- 보상(reward)을 통한 학습
- 에이전트는 현재의 환경에 대해 행동을 결정하여 실행하고, 이에 따라 변화된 환경과 함께 보상을 받음



⇒ 일련의 행동에 따른 보상을
최대화 하는 정책을 개발



귀납적 학습

1. 지식의 일반화

■ 기계적 학습

- 암기식 학습(rote learning)
- 문제와 풀이를 그대로 저장하여 동일한 상황에서 다시 사용할 수 있도록 함

입력: (X_1, X_2, \dots, X_n)  출력: (Y_1, Y_2, \dots, Y_p)

지식

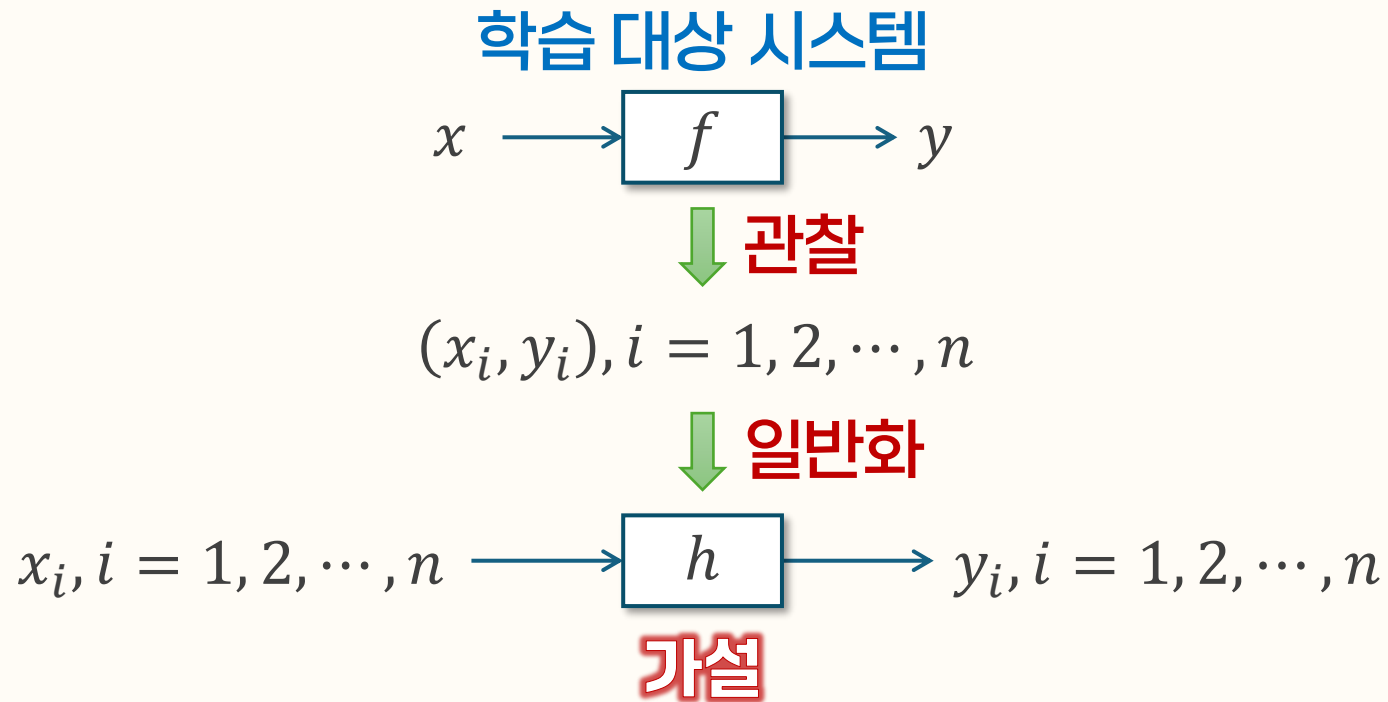
$$\begin{array}{c} \vdots \\ [(X_1, X_2, \dots, X_n), (Y_1, Y_2, \dots, Y_p)] \\ \vdots \end{array}$$

 일반화를 통한 효율적인 학습이 필요함

2. 귀납적 학습의 개념

■ 귀납적 추론과 학습

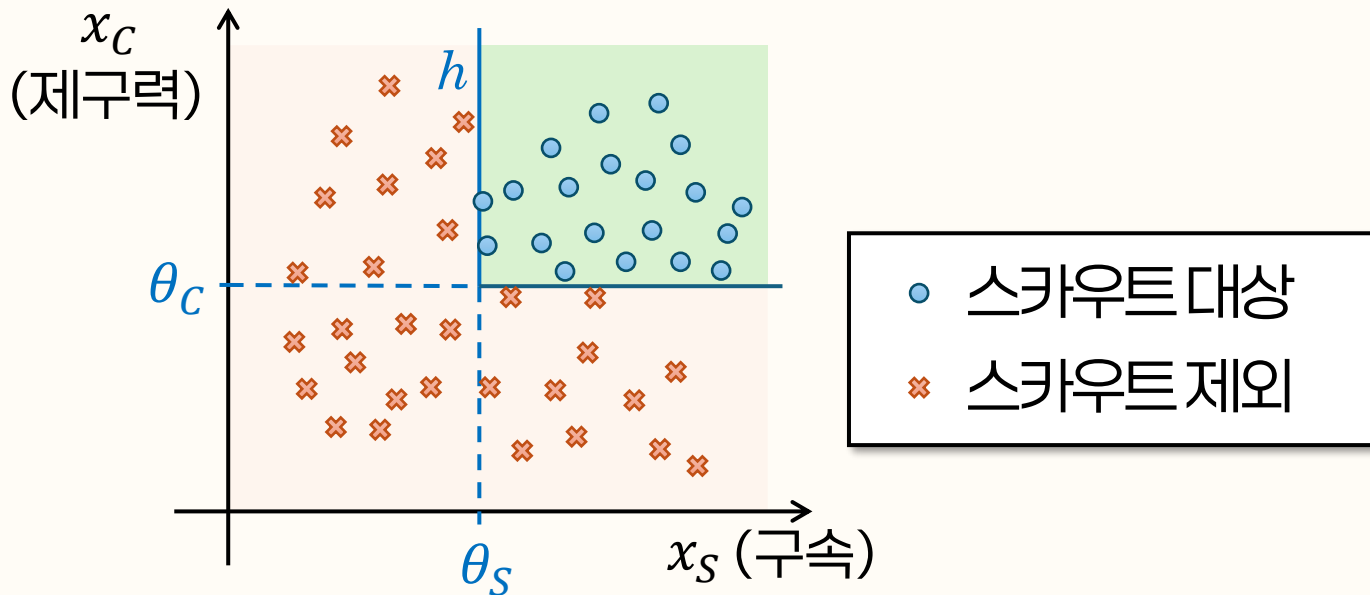
- 귀납적 추론: 반복적인 관찰이나 경험을 기반으로 해서 일반화된 논리를 이끌어내는 추론방법



2. 귀납적 학습의 개념

예 : 스카우터의 우수 투수 판별 규칙

IF $x_{\text{구속}} > \theta_S$ AND $x_{\text{제구력}} > \theta_C$ THEN 스카우트(x)



⇒ 귀납적 학습으로 일반화된 규칙을 학습함으로써 학습표본 집합과 유사한 입력에 대해 적절한 판단을 내릴 수 있음

2. 귀납적 학습의 개념

■ 예 : 스카우터의 우수 투수 판별 규칙

IF $x.\text{구속} > \theta_s$ AND $x.\text{제구력} > \theta_c$ THEN 스카우트(x)



학습에 사용되는 예는 실제로 발생할 수 있는 모든 경우 중에서 일부에 해당됨

- ⇒ 학습에 사용되는 예의 집합은 다양하게 구성될 수 있음
- ⇒ 사용된 학습표본 집합에 따라 동일한 입력에 대해 다른 판단을 내리는 경우도 발생할 수 있음
- ⇒ 학습 과정에서 제시되지 않은 입력에 대해서는 잘못된 판단을 내릴 수도 있음

2. 귀납적 학습의 개념

이진 분류기의 평가 기준

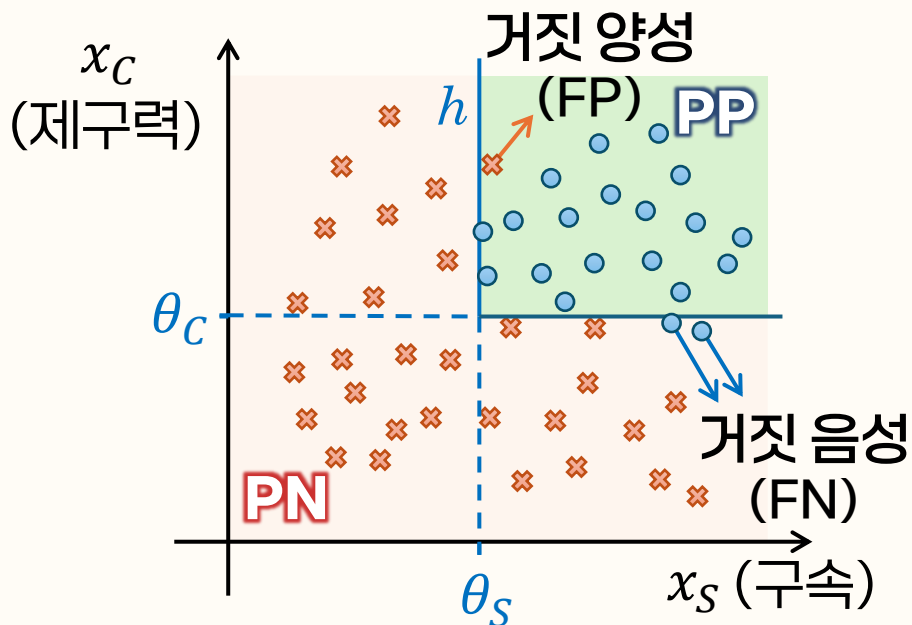
- 이진 분류기의 분할표

레이블 \ 분류기의 예측	양성 (Predicted Positive, PP)	음성 (Predicted Negative, PN)
양성 (Positive, P)	참 양성 (True Positive, TP)	거짓 음성 (False Negative, FN)
음성 (Negative, N)	거짓 양성 (False Positive, FP)	참 음성 (True Negative, TN)

2. 귀납적 학습의 개념

이진 분류기의 평가 기준

IF $x.\text{구속} > \theta_s$ AND $x.\text{제구력} > \theta_c$ THEN 스카우트(x)

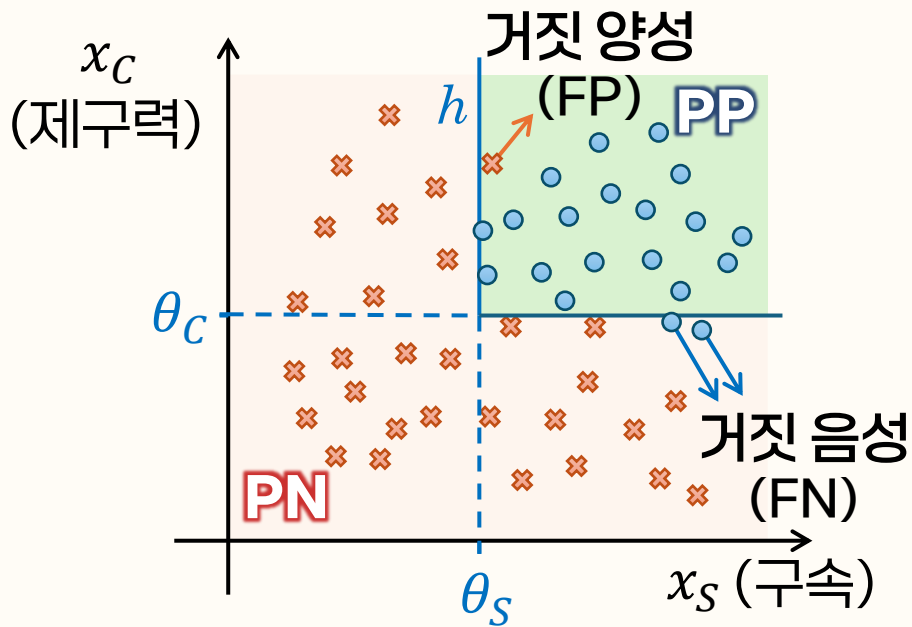


정밀도: $precision = \frac{TP}{TP + FP} = \frac{18}{19} \approx 0.95$

2. 귀납적 학습의 개념

이진 분류기의 평가 기준

IF $x.\text{구속} > \theta_s$ AND $x.\text{제구력} > \theta_c$ THEN 스카우트(x)

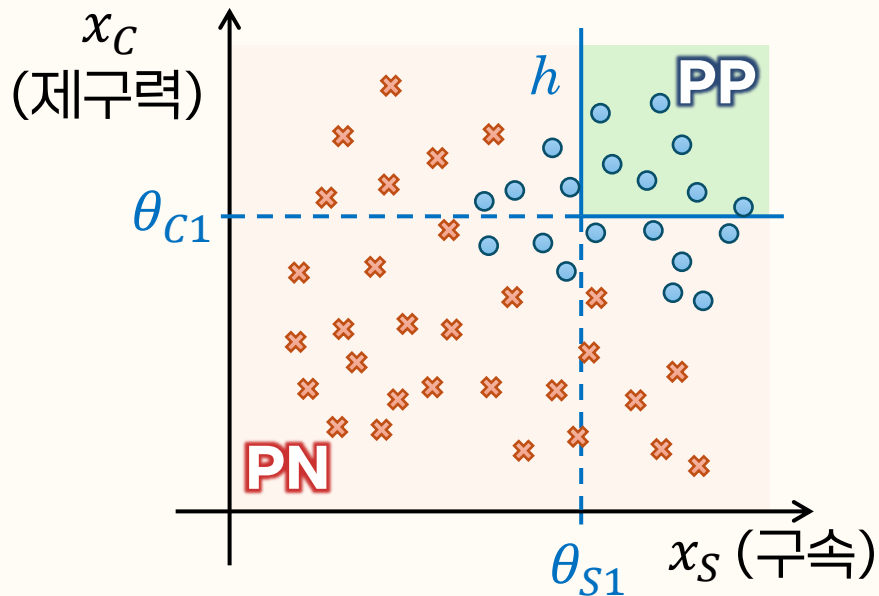


재현율: $recall = \frac{TP}{TP + FN} = \frac{18}{20} = 0.9$

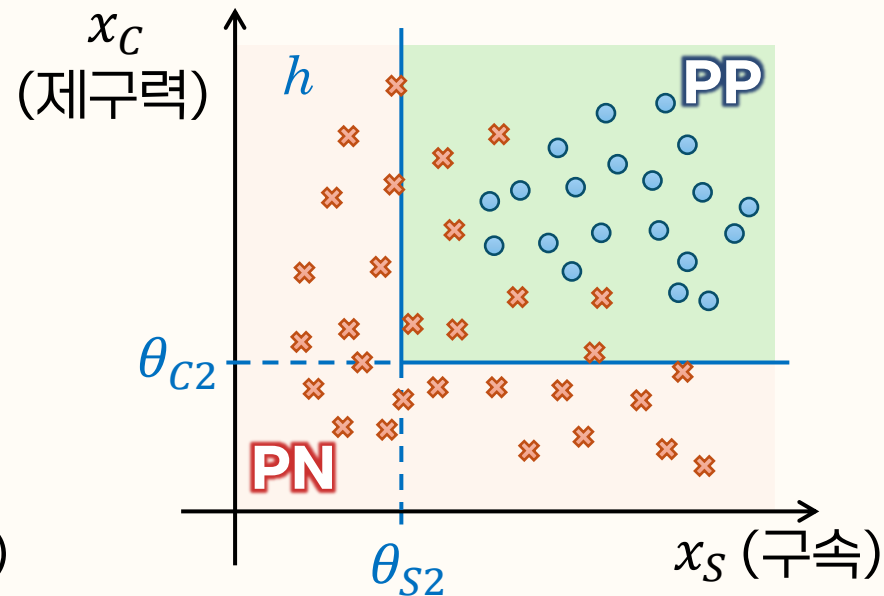
2. 귀납적 학습의 개념

이진 분류기의 평가 기준

IF $x_{\text{구속}} > \theta_s$ AND $x_{\text{제구력}} > \theta_c$ THEN 스카우트(x)



⇒ $precision = 1.0$
 $recall = 0.35$



⇒ $recall = 1.0$
 $precision = 0.71$

2. 귀납적 학습의 개념

이진 분류기의 평가 기준

 F_1 점수: 정밀도와 재현율의 조화평균

$$\begin{aligned} F_1 &= 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \\ &= \frac{2TP}{2TP + FP + FN} \end{aligned}$$

 정확도: 전체 데이터 중 분류기가 올바르게 분류한 데이터의 비율

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

2. 귀납적 학습의 개념

■ 학습 데이터와 학습의 질

- 학습 데이터 집합의 결함은 학습 품질과 밀접한 관계가 있음

★ 학습 예의 결함 유형

- 1 부정확한 입력 속성값이 존재하는 경우
- 2 학습 예를 잘못 분류한 경우
- 3 학습에 필요한 주요 속성이 빠진 경우



결정트리 학습

1. 결정트리

■ 결정트리(decision tree)란?

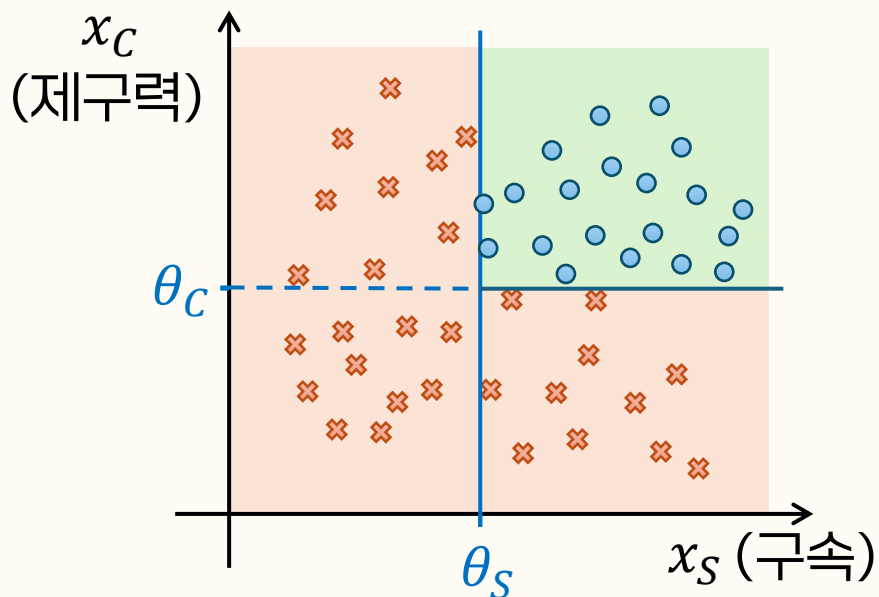
- 분할정복 방식으로 특징공간을 분할함으로써 입력된 대상을 분류(분류트리, classification tree)하거나 회귀분석(회귀트리, regression tree)을 하기 위한 트리

■ 결정트리의 노드

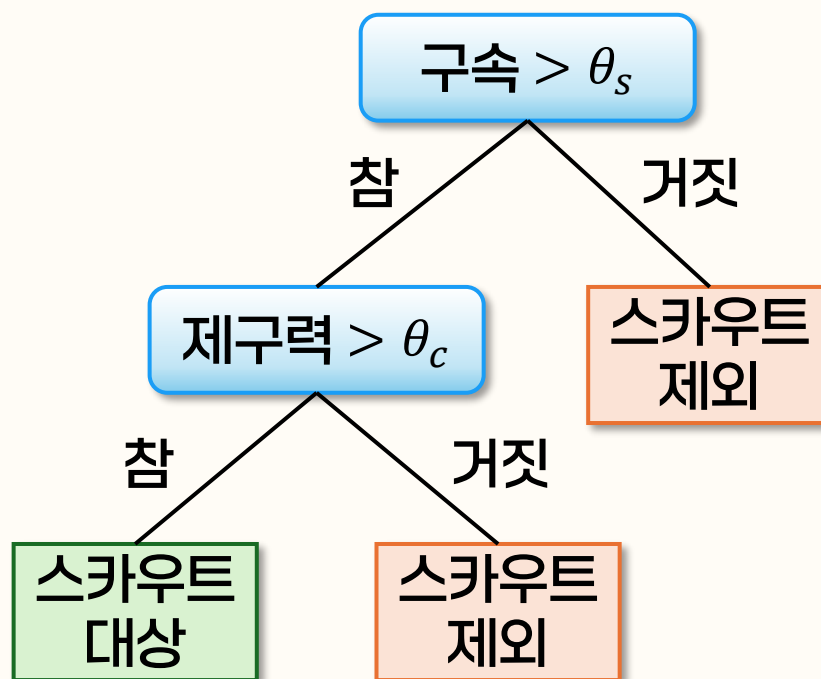
- 루트 및 내부노드(결정노드, decision node): 공간을 분할하는 조건을 판단하는 함수를 갖고 있음
- 잎노드(leaf node): 최종적인 분할영역을 나타내는 노드로, 입력에 대한 출력 값을 갖고 있음

1. 결정트리

결정트리(decision tree)란?



- 스카우트대상
- ✕ 스카우트제외



2. 결정트리의 학습

■ 결정트리 학습의 개념

- 제시된 학습표본 집합을 이용하여 지도학습 방식으로 학습
- 각각의 노드에서 그 노드에 분류된 학습표본 부분집합에 대해 불순도 검사(impurity test) 시행
 - ⇒ 불순도가 높으면 적절한 검사를 통해 이를 분할할 수 있는 결정노드를 생성
 - ⇒ 불순도가 0(또는 정해진 임계치 미만)이면 해당 클래스(또는 다수의 표본이 속하는 클래스)를 출력하는 잎 노드를 생성

2. 결정트리의 학습

❏ 불순도 검사 방법

- 엔트로피(entropy), 지니 불순도(Gini impurity) 등 사용

★ 엔트로피를 이용한 불순도 ϕ_E 의 계산

$$\phi_E = - \sum_{i=1}^N p_i \log_2 p_i$$

N : 분류하고자 하는 클래스 수

$$p_i = m_i / m$$

m : 노드에 속한 학습표본 부분집합의 표본 수

m_i : i 번째 클래스에 속하는 표본 수

2. 결정트리의 학습

❏ 불순도 검사 방법

- 엔트로피(entropy), 지니 불순도(Gini impurity) 등 사용

★ 지니 불순도 ϕ_G 계산

$$\phi_G = 1 - \sum_{i=1}^N p_i^2$$

N : 분류하고자 하는 클래스 수

$$p_i = m_i / m$$

m : 노드에 속한 학습표본 부분집합의 표본 수

m_i : i 번째 클래스에 속하는 표본 수

2. 결정트리의 학습

■ 노드 분할

- 분할 결과의 불순도가 가장 낮아지게 하는 속성을 기준으로 분할

$$C(j) = \sum_{k=1}^K \frac{m_k^{(j)}}{m} \phi_k^{(j)}$$

$C(j)$: j 번째 속성을 기준으로 분할한 결과의 불순도

m : 분할할 노드에 분류된 학습표본 수

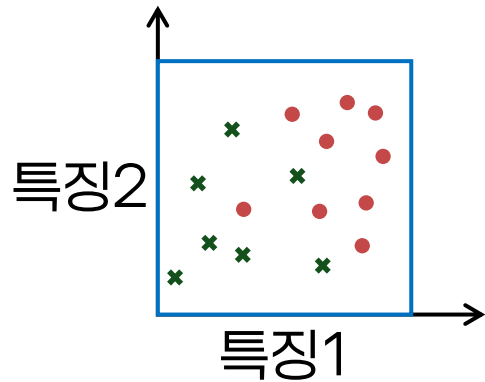
$m_k^{(j)}$: k 번째 자식노드에 분류되는 학습표본 수

$\phi_k^{(j)}$: k 번째 자식노드의 불순도

2. 결정트리의 학습

■ 노드 분할

예 지니 불순도를 이용한 노드 분할



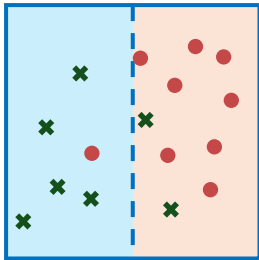
$$\begin{aligned}
 \Rightarrow \phi &= 1 - \left\{ \left(\frac{7}{16} \right)^2 + \left(\frac{9}{16} \right)^2 \right\} \\
 &= 1 - \frac{130}{256} \\
 &\approx 0.492
 \end{aligned}$$

2. 결정트리의 학습

■ 노드 분할

예 지니 불순도를 이용한 노드 분할

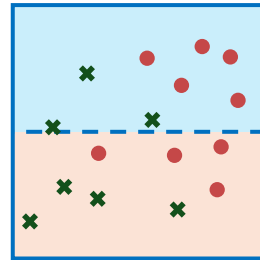
$$\phi_1^{(1)} = 0.278$$



$$\phi_2^{(1)} = 0.32$$

$$\begin{aligned} \Rightarrow C(1) &= \frac{6}{16} \cdot 0.278 + \frac{10}{16} \cdot 0.32 \\ &= 0.304 \end{aligned}$$

$$\phi_1^{(2)} = 0.469$$



$$\phi_2^{(2)} = 0.5$$

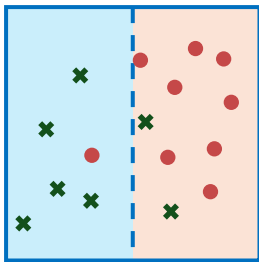
$$\begin{aligned} \Rightarrow C(2) &= \frac{8}{16} \cdot 0.469 + \frac{8}{16} \cdot 0.5 \\ &= 0.485 \end{aligned}$$

2. 결정트리의 학습

■ 노드 분할

예 지니 불순도를 이용한 노드 분할

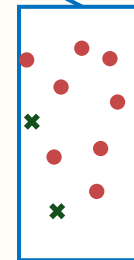
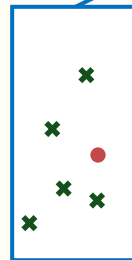
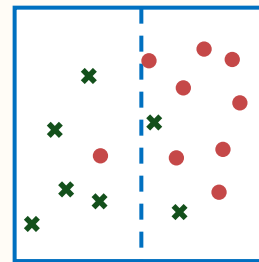
$$\phi_1^{(1)} = 0.278$$



$$\phi_2^{(1)} = 0.32$$

$$\begin{aligned} \Rightarrow C(1) &= \frac{6}{16} \cdot 0.278 + \frac{10}{16} \cdot 0.32 \\ &= 0.304 \end{aligned}$$

비단말노드



정리하기

- ✔ 머신러닝이란 인공지능 시스템이 지능적 행동능력을 갖기 위해서 외부 환경으로부터의 정보를 이용하여 시스템 내부에 지식을 형성하고 저장하는 과정이다.
- ✔ 학습을 위해 제공되는 정보의 유형에 따라 지도학습, 비지도학습, 강화학습 등 여러 가지 학습 방법이 존재하며, 이를 변형한 준지도학습, 자기지도학습 등의 학습 방법이 있다.
- ✔ 귀납적 추론은 학습을 위한 기본적 토대로서, 반복적인 관찰이나 경험을 기반으로 해서 일반화된 논리를 이끌어 내는 추론을 한다.

정리하기

- ✓ 결정트리는 노드의 불순도를 구하여 불순도가 큰 노드를 적절한 특징을 기준으로 분할하는 과정을 통해 형성하며, 지도학습 방식을 따른다.
- ✓ 노드의 불순도 측정은 엔트로피나 지니 불순도 등을 사용할 수 있다.

11강

다음시간안내 ▶▶▶

머신러닝(2)