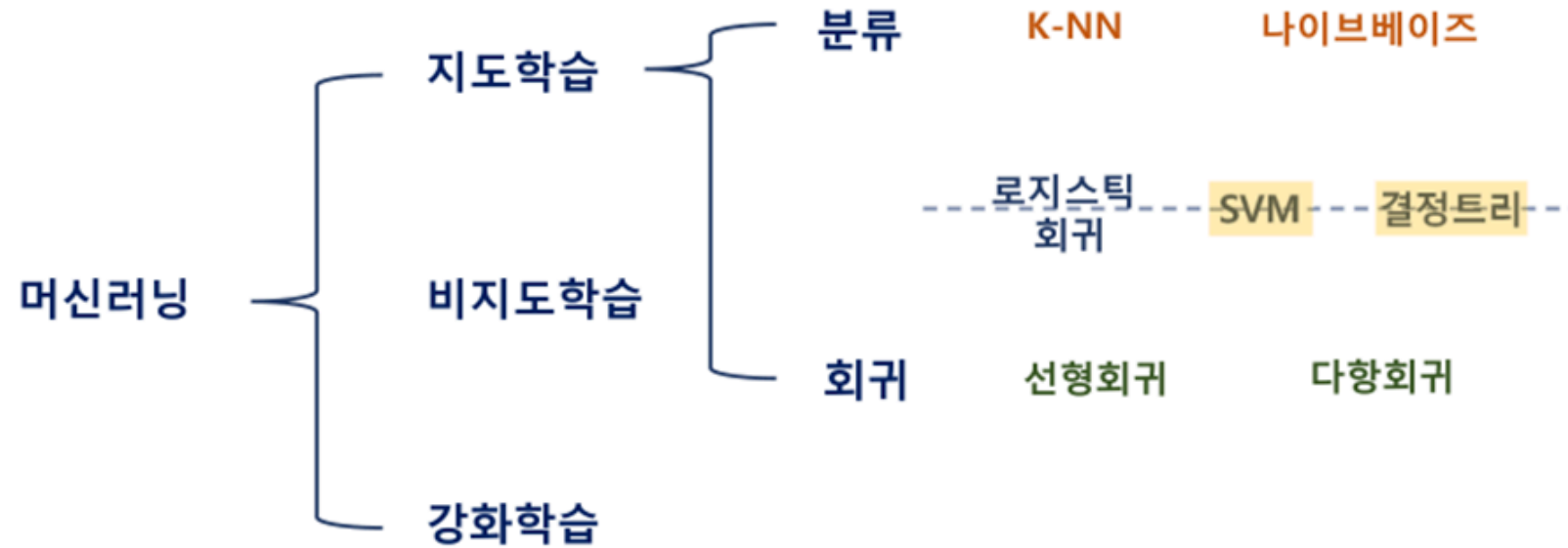


- ✓ SVM과 결정트리(decision tree)

7강. SVM과 결정트리

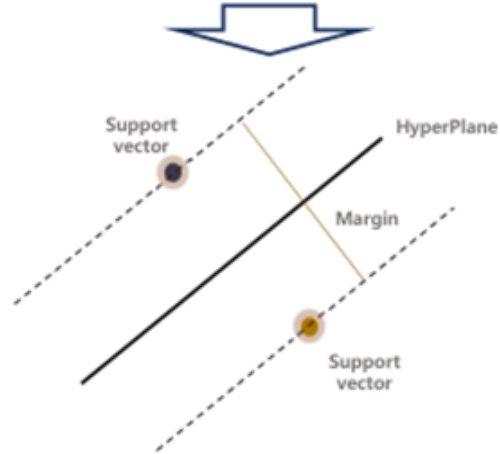
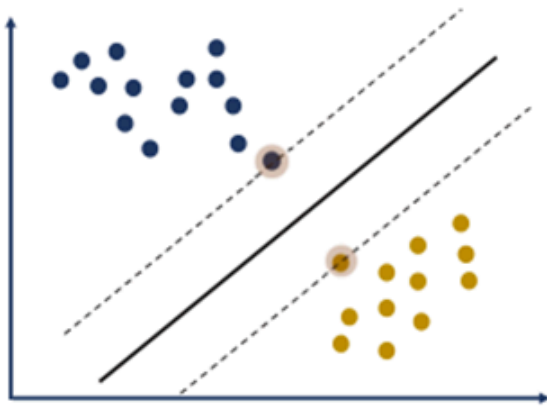
- SVM(Support Vector Machine)
- 결정트리(Decision Tree)
- 지니계수와 엔트로피계수

지도 학습의 종류



선형이나 비선형 분류, 회귀 등에서 활용할 수 있는 다목적 머신러닝 모델

SVM(Support Vector Machine)



Hyperplane

데이터를 구분하는 기준이 되는 경계

Support Vector

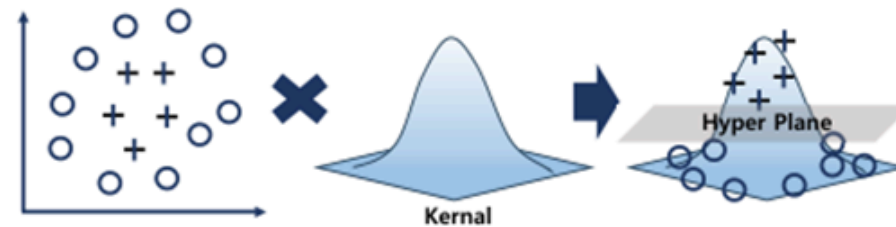
클래스를 나누는 Hyper Plane과 가장 가까운 위치의 샘플

Margin

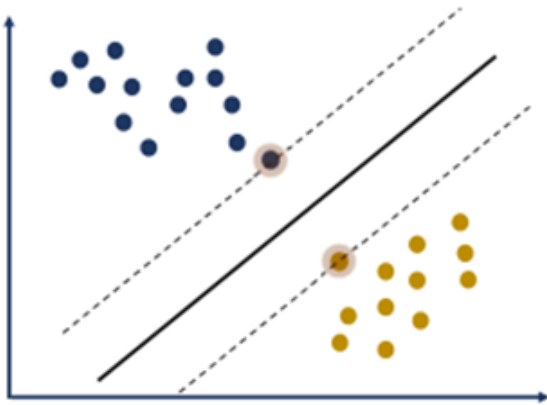
Hyper Plane과 Support Vector 사이의 거리

Kernal 함수

저차원의 데이터를 고차원으로 변경하는 함수

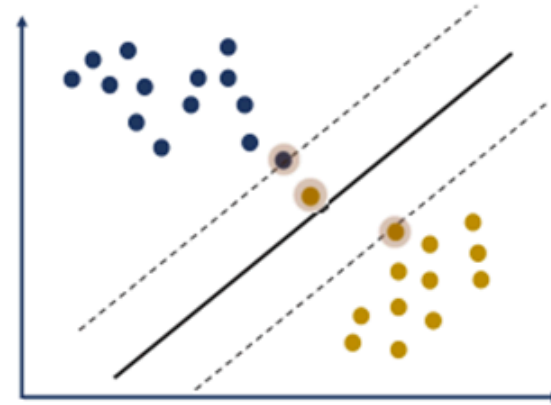


SVM(Support Vector Machine)



하드 마진 분류

- 모든 샘플이 마진 바깥쪽에 분류
- 이상치에 민감



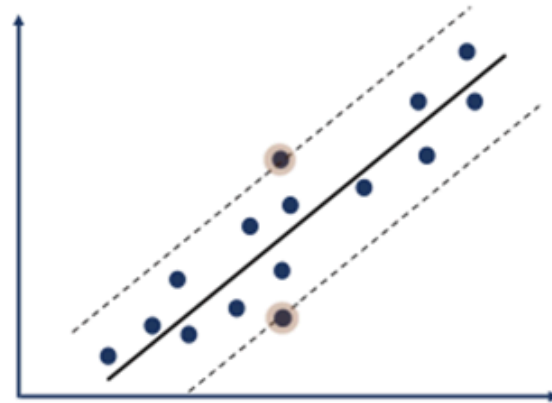
소프트 마진 분류

- 마진 내 어느 정도의 오류를 감안
- 하이퍼파라미터로 마진 넓이 조절 가능

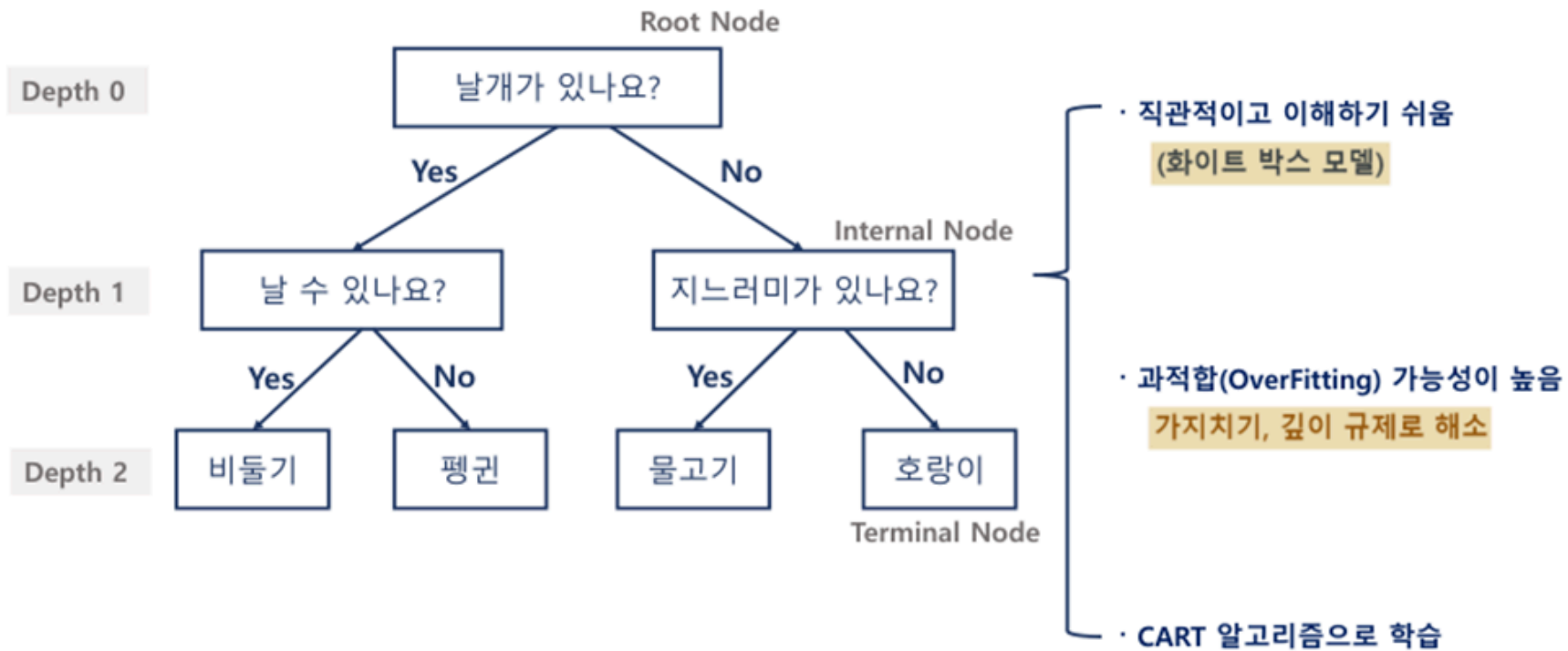
SVM(Support Vector Machine)

회귀 모델에서의 SVM

마진 오류안에 가능한 많은 샘플이 들어가도록 학습



결정 트리(Decision Tree)

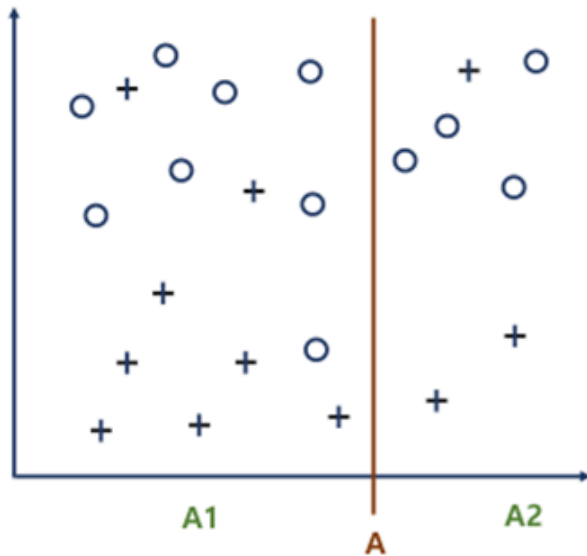


CART 알고리즘

Classification and Regression Tree

불순도가 낮아지는 방향으로 나무를 분리하는 기법

$$\text{지니계수 } G = 1 - \sum_{i=1}^m P_i^2$$



A를 기준으로 지니계수 계산

$$\cdot G_{A1} = 1 - \left(\frac{8}{16}\right)^2 - \left(\frac{8}{16}\right)^2 = 0.5$$

$$\cdot G_{A2} = 1 - \left(\frac{4}{7}\right)^2 - \left(\frac{3}{7}\right)^2 = 0.49$$

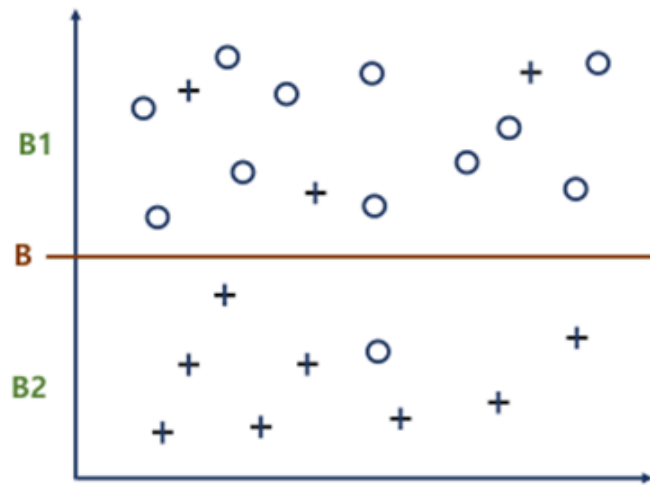
$$\begin{aligned} \cdot G_A &= \left(\frac{16}{23}\right)G_{A1} + \left(\frac{7}{23}\right)G_{A2} \\ &= \left(\frac{16}{23}\right)0.5 + \left(\frac{7}{23}\right)0.49 = 0.497 \end{aligned}$$

CART 알고리즘

Classification and Regression Tree

불순도가 낮아지는 방향으로 나무를 분리하는 기법

$$\text{지니계수 } G = 1 - \sum_{i=1}^m P_i^2$$



B를 기준으로 지니계수 계산

$$\cdot G_{B1} = 1 - \left(\frac{11}{14}\right)^2 - \left(\frac{3}{14}\right)^2 = 0.34$$

$$\cdot G_{B2} = 1 - \left(\frac{1}{9}\right)^2 - \left(\frac{8}{9}\right)^2 = 0.2$$

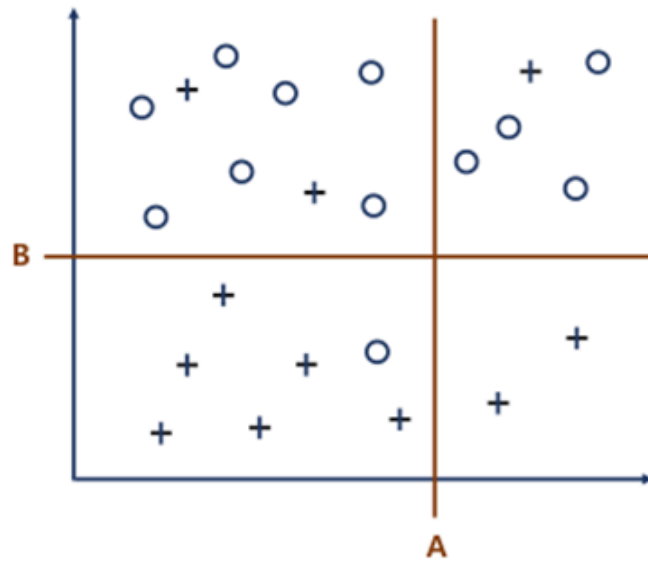
$$\begin{aligned} \cdot G_B &= \left(\frac{14}{23}\right)G_{B1} + \left(\frac{9}{23}\right)G_{B2} \\ &= \left(\frac{14}{23}\right)0.34 + \left(\frac{9}{23}\right)0.2 = 0.28 \end{aligned}$$

CART 알고리즘

Classification and Regression Tree

불순도가 낮아지는 방향으로 나무를 분리하는 기법

$$\text{지니계수 } G = 1 - \sum_{i=1}^m P_i^2$$



최초 불순도 = 1

$$G_A = 0.497$$

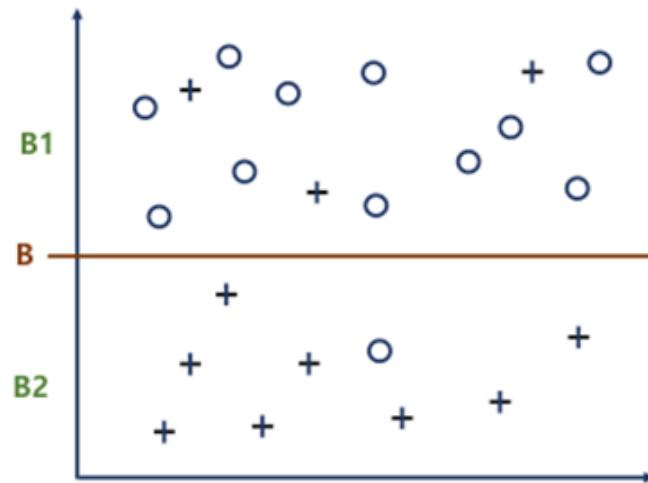
$$G_B = 0.28$$

CART 알고리즘

Classification and Regression Tree

불순도가 낮아지는 방향으로 나무를 분리하는 기법

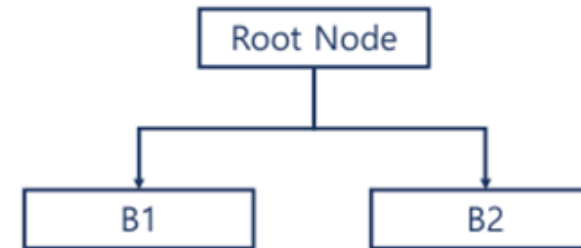
$$\text{지니계수 } G = 1 - \sum_{i=1}^m p_i^2$$



최초 불순도 = 1

$$G_A = 0.497$$

$$G_B = 0.28$$



■ 지니계수와 엔트로피계수