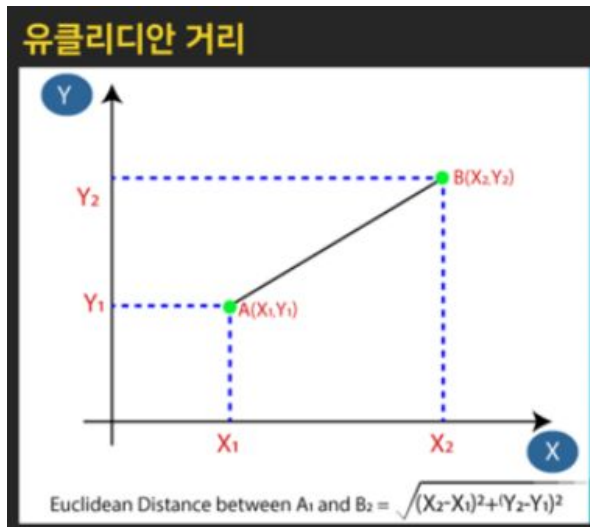
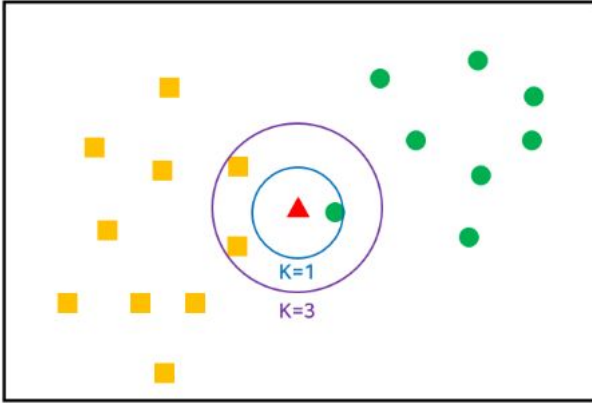


KNN

- K-최근접 이웃 알고리즘
- 거리기반 분류모듈
- 거리가 가까운 K개의 다른 데이터의 레이블을 참조하여 분류
- 거리 측정시 유클리디안 거리 계산법 사용





K값에 따라 분류가 달라질 수 있음

장점

알고리즘이 단순하다

훈련단계가 빠르다

수치기반 데이터 분류 작업의 성능이 좋다

단점

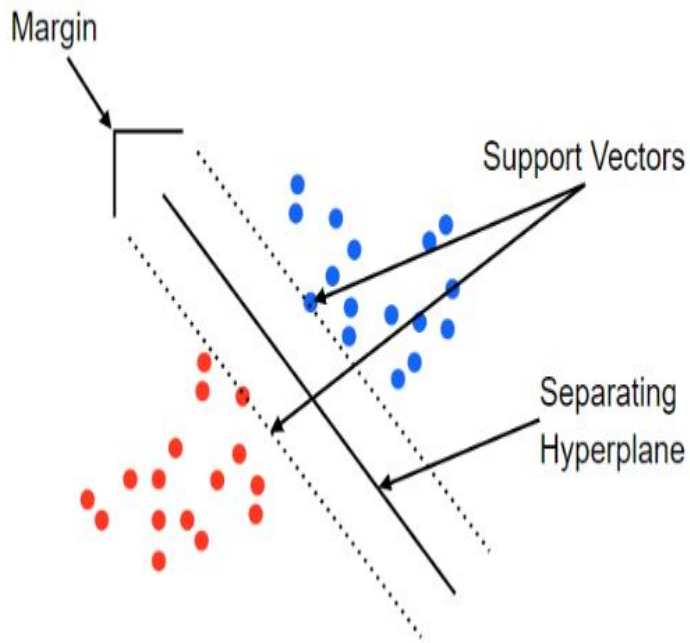
특징과 클래스간의 관계이해가 어렵다

적절한 K의 선택이 필요하다

데이터가 많아지면 분류 단계가 느리다

SVM

- 서포트벡터머신(Support Vector Machine)
- 데이터 분리를 위한 선이 필요
- Margin을 이용한 분리
- 직선을 초평면(hyperplane) 이라함



Hyperplane(초평면):

데이터 분류 칸막이

Support Vector(서포트벡터) :

초평면에 가장 가까운 데이터 들

Margin(마진) :

양쪽 서포트벡터와 초평면간의거리

- 장점

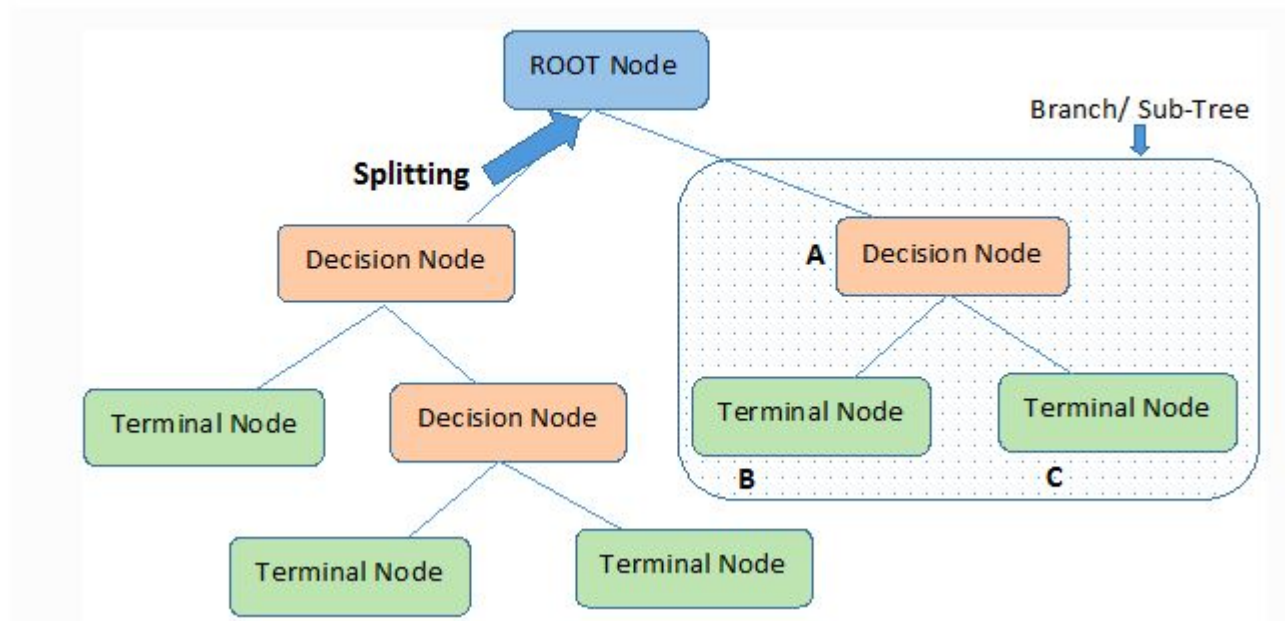
- 분류나 예측에 사용 가능
- 과적합 정도가 덜함
- 오류 데이터 영향이 적음

- 단점

- 여러 개의 조합 테스트가 필요
- 학습 속도가 느림

Decision Tree

- 데이터에 있는 규칙을 통해 데이터셋을 분류/예측하는 지도학습(supervised) 모델



노드 : 네모칸

결정노드(Decision Node) : 규칙조건

리프 노드(leaf node) : 마지막노드

terminal node라고도 함

에지(edge) : 질문의 답과 다음 질문을 연결

- 장점
 - 알고리즘이 쉽다
 - 시각화 표현에 유리
 - 각 피처의 정규화필요 없음
- 단점
 - 과적합 발생 확률이 높다