

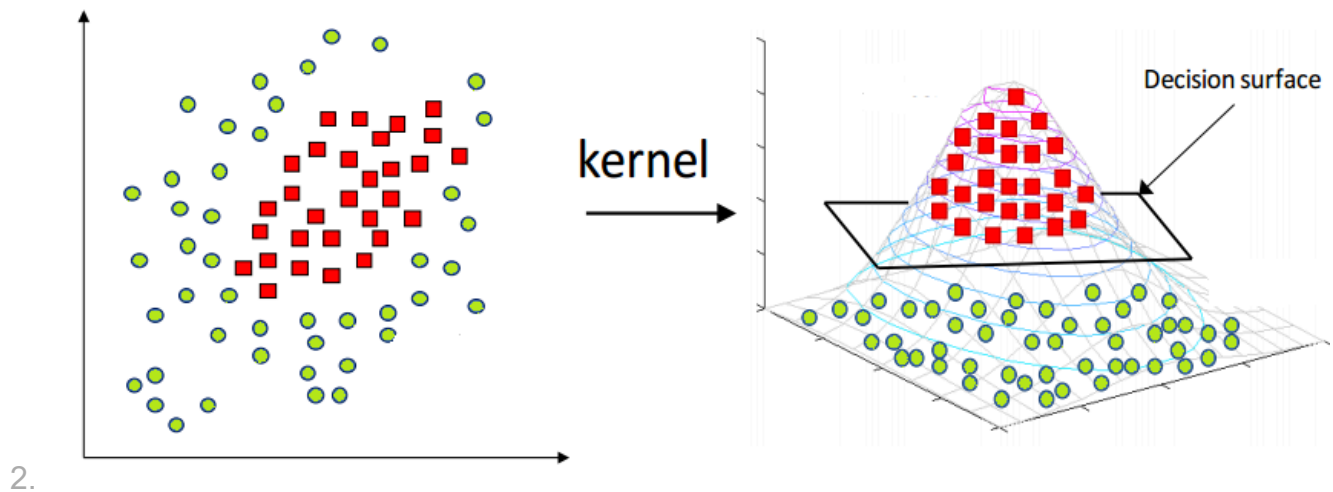
Notions

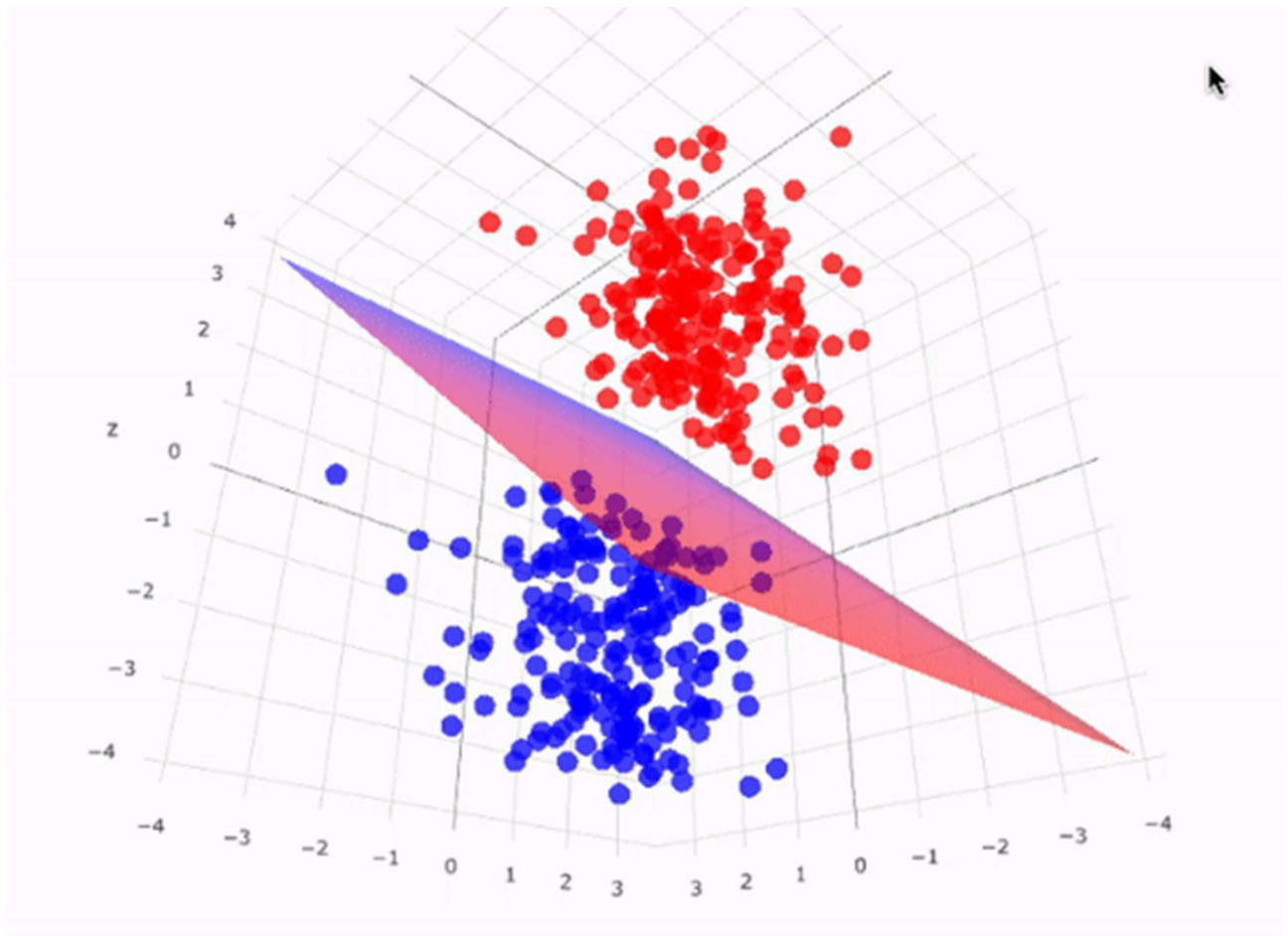
Classification Algorithm

SVM (Support Vector Machine)

<https://kr.mathworks.com/discovery/support-vector-machine.html>

1. 목표: 두 class를 최대한 잘 분류하는 조평면을 찾는 것
 1. 조평면: n 차원 공간 속 $n-1$ 차원 (3차원 공간 속 평면)
 2. 두 class 사이에 가장 큰 margin을 갖는 조평면으로 정의





3.

RF (Random Forest)

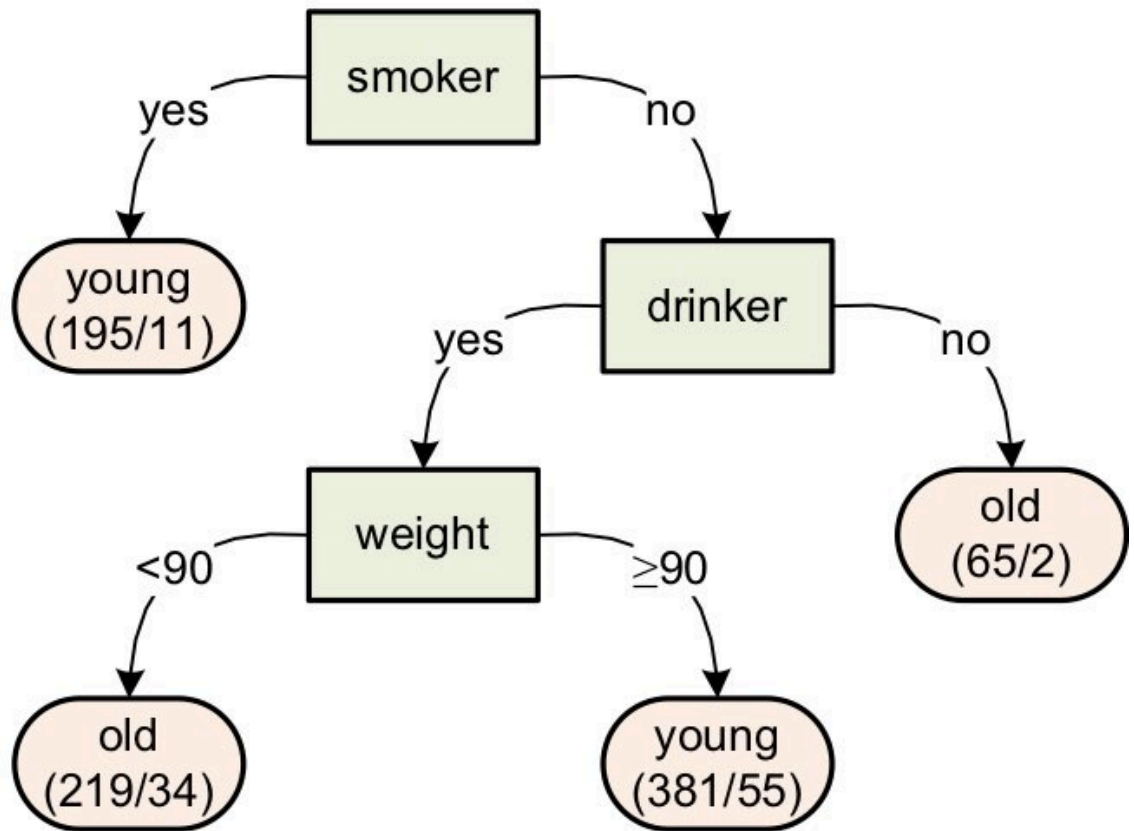
<https://www.ibm.com/kr-ko/topics/random-forest>

의사 결정 트리의 포레스트를 만드는 것 (DT → Ensemble)

1. DT (Decision Tree)

<https://bkshin.tistory.com/entry/%EB%A8%B8%EC%8B%A0%EB%9F%AC%EB%8B%9D-4-%EA%B2%B0%EC%A0%95-%ED%8A%B8%EB%A6%ACDecision-Tree>

1. 특정 기준에 따라 데이터를 구분하는 트리 형태의 모델
 1. Node: 각 질문 및 정답 상자
2. Tree 구조를 활용해 Entropy를 최소화하는 방향으로 데이터 분류 및 예측
3. Impurity (불순도): 해당 범주 안에 서로 다른 데이터가 섞여있는 정도
4. Entropy: Impurity를 수치화한 척도
 1. 높은 Entropy, 높은 Impurity
 2. Entropy == 1: Impurity가 최대이며, 해당 범주 안에 데이터가 정확히 반반 존재



3.

2. Ensemble

<https://medium.com/dawn-cau/%EB%A8%B8%EC%8B%A0%EB%9F%AC%EB%8B%9D-%EC%95%99%EC%83%81%EB%B8%94-%ED%95%99%EC%8A%B5-%EC%9D%B4%EB%9E%80-cf1fcb97f9d0>

1. SVM 등의 Classifier를 조합해 최적의 모델로 일반화하는 방법
2. Voting, Bagging, Boosting, Stacking
3. Bagging (Bootstrap Aggregation)
 1. 샘플을 여러 번 뽑아서 각 모델 학습 및 집계
 2. 복원 추출한 데이터로 모델 학습
 3. 서로 다른 데이터를 통해 동일한 알고리즘의 분류기 사용

Recommndation Algorithm

- Collaborative Filtering
 - Memory Based Approach
 - User-based Collaborative Filtering
 - Item-based Collaborative Filtering
 - Model Based Approach (ML based)
- Contents-based Filtering

Collaborative Filtering

<https://www.blossominkyung.com/recommendersystem/collaborative-filtering>

<https://kmhana.tistory.com/31>

- 사용자의 경험과 행동 방식에 의존
- 사용자와 아이템 간의 상호 상관 관계를 분석해 새로운 사용자-아이템 관계 탐색

Memory Based CF

- 사용자 간의, 아이템 간의 관계 계산에 중점
 - 아직 평가하지 않은 아이템 예측
1. User-based Collaborative Filtering
 2. 유사한 아이템에 대해 비슷한 점수를 매긴 사용자 탐색
 3. 해당 사용자가 상호 작용하지 않은 아이템에 대한 아이템의 점수 예측
 2. Item-based Collaborative Filtering (선택)
 1. 동일한 사용자의 이웃 아이템 점수를 기반으로 해당 아이템에 대한 사용자의 선호도 평가
 2. 이미 상호 작용한 사용자를 대상으로 하는 아이템과 유사한 아이템 탐색

Word Embedding

Word2Vec

<https://wikidocs.net/22660>

1. 단어 벡터 간의 유사도 연산을 위해 단어의 의미 수치화
2. sliding window
 1. window를 움직여 주변/중심 단어를 변경해 데이터셋 생성
 1. window size: 중심 단어 예측을 위해 참고할 중심 단어 앞뒤 N개의 단어
 2. 중심 단어 예측을 위해 참고하는 주변 단어의 수: 2N
 2. input: 주변 단어의 one-hot vector
 3. label: 중간 단어의 one-hot vector
- 3.