

Introduction

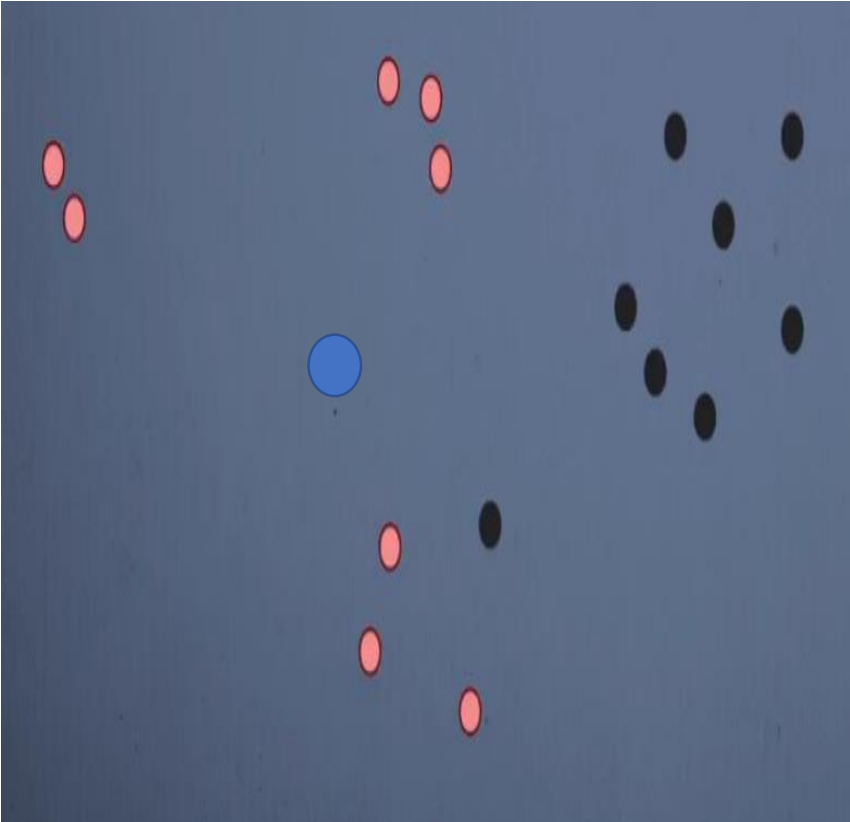
Machine Learning

- Program의 performance를 given data, previous results 혹은 experiences를 이용해서 향상시키는 것
- 주어진 data 혹은 이전의 결과 혹은 이전의 경험들을 활용해서, 그것을 기반으로 하여 자동적으로 program의 성능을 향상시킬 것이라는 기법
- Data에 존재하는 pattern 찾기
 - Valid : Data에 존재하는 올바른 pattern을 찾았는가?
 - Novel : 새로운 pattern을 찾았는가?
 - Useful : 성능을 높일 수 있는가?(유용한가?)
 - Understandable : 의미가 무엇인가?

ML Problem

- Classification
- Regression
- Recommendation
- Clustering
- Dimension Reduction
- 기타 등등..

Classification

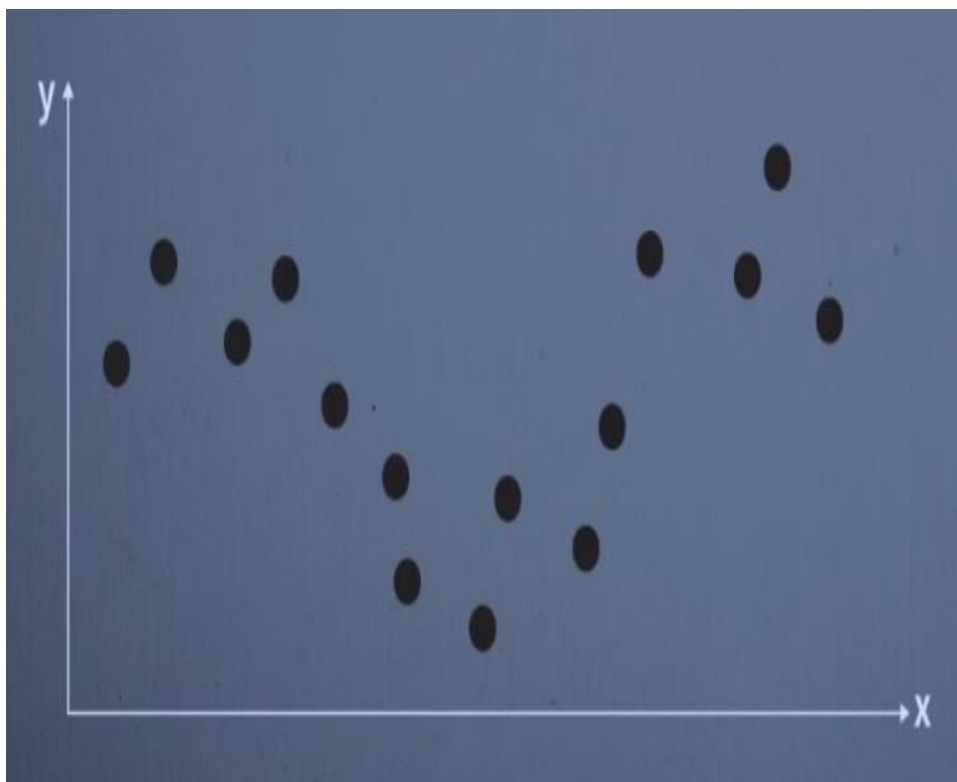


빨간색 : 양품

검은색 : 불량품

Unknown data에 대하여 불량품인지 양품인지
"예측" 하는 문제

Regression



X: 입력

Y: 출력

과거의 data에서 x일 때 y값을 수집

새로운 data가 들어왔을 때(x) 어떤 결과 값(y) 인지
"예측" 하는 문제

주로 real number 예측(classification과 거의 동일)

Recommendation

	I_1	I_2	I_3	I_4	I_5
U_1	?	3	?	5	4
U_2	2	?	2	3	3
U_3	?	2	4	?	?
U_4	2	3	3	2	5
U_5	1	?	1	?	2

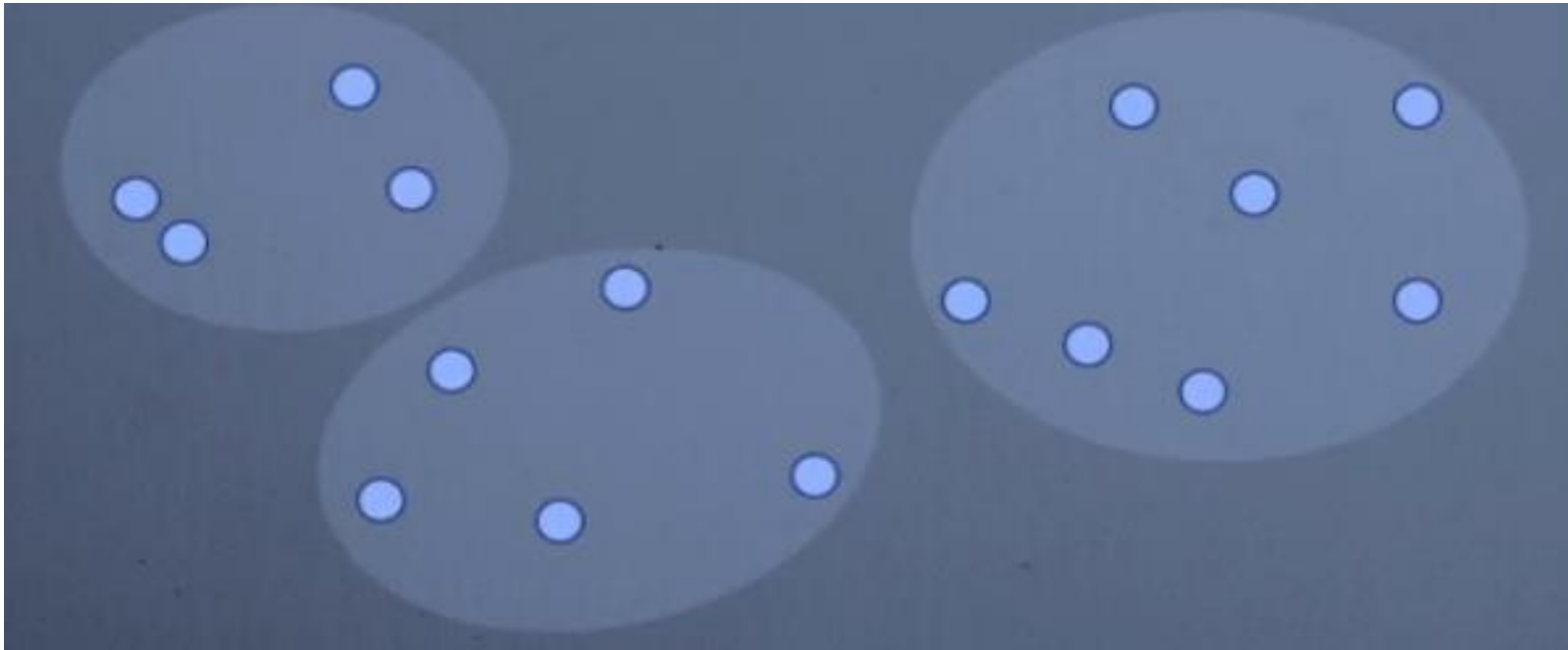
I : 영화

U : 관람자

User가 어떤 item을 사용했고, 어떤 item을 사용하지 않았는지 알았을 때 다음에 유저가 어떤 item을 사용할 것인지 "예측" 하는 문제

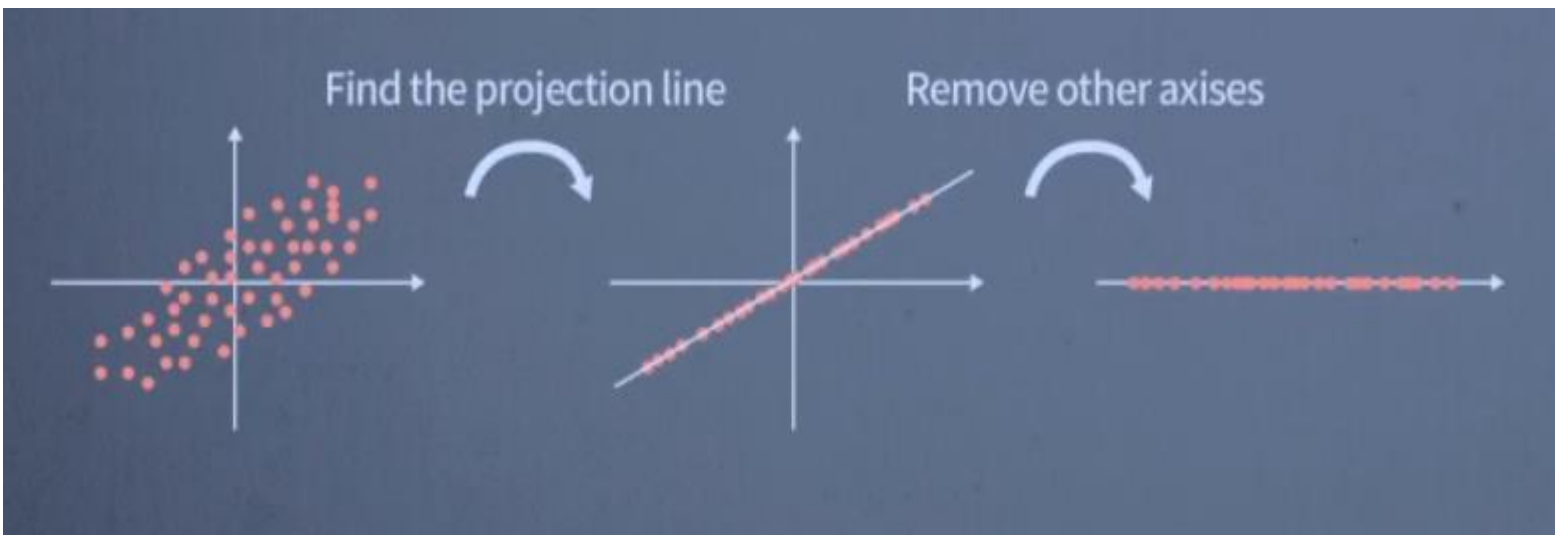
Clustering

- Data 자체에 대한 궁금증을 해석하기 위한 분석기술
- 평균, 분산 등 계산
- 주어진 data에 대한 유사한 그룹 찾아낼 때 쓰는 것



Dimension Reduction

- Data 차원을 줄여서 내가 처리해야되는 정보의 양을 줄이는 것



ML Methods

- Supervised Learning
 - Classification, Regression
 - 누군가가 예측이 맞았는지, 틀렸는지 알려줘야 함
- Semi-supervised Learning
 - Data 맞았는지 틀렸는지 알려주기 애매한 경우
- Unsupervised Learning
 - Data 분석
 - Clustering, Dimension Reduction
- Reinforcement Learning
 - 어떤 전략, 행동을 했을 때 Reward를 최대화하는 방법 찾기