

전통적 방식 computing

: 문제 \rightarrow possible, but practically impossible

\rightarrow 절차적으로 표현하기가 힘든 상황

대안 방식 computing

인공지능 : 탐색을 통한 문제해결, 지식표현과 논리적 추론, 학습

기계학습(machine learning):

- Supervised Learning(지도학습)
- Unsupervised Learning(비지도학습)
- Reinforcement Learning(강화학습)

지도학습, 비지도 학습 : 과거의 경험인 데이터를 사용한다.

강화학습 : trial and error \rightarrow 경험에 의한 학습

지도 \rightarrow 분류(class) \rightarrow Decision Tree

비지도 \rightarrow 군집 \rightarrow K-mean 알고리즘

\rightarrow 연관 \rightarrow Apriori 알고리즘

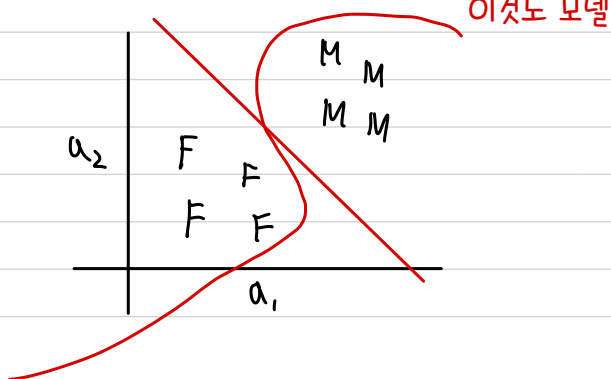
강화학습

-Decision Tree-

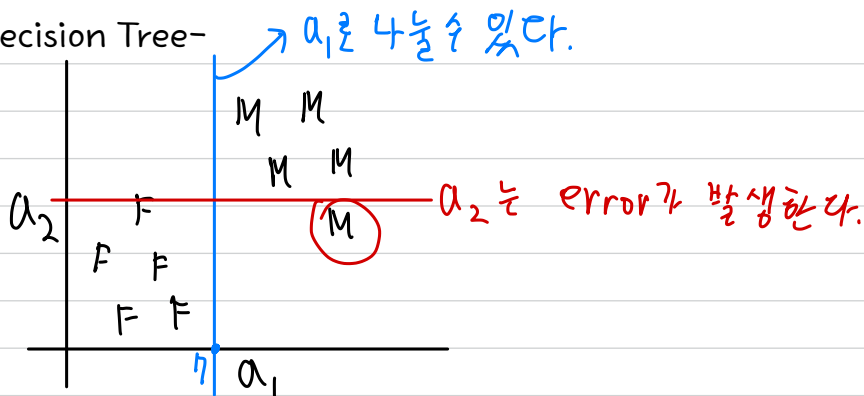
데이터

feature attribute		class
a_1	a_2	
20	25	M
55	65	M
;	;	F
;	;	;

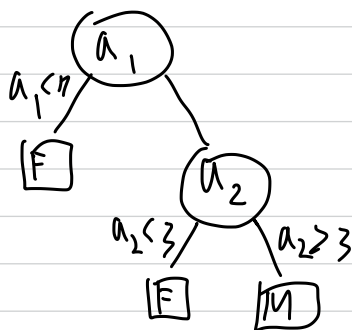
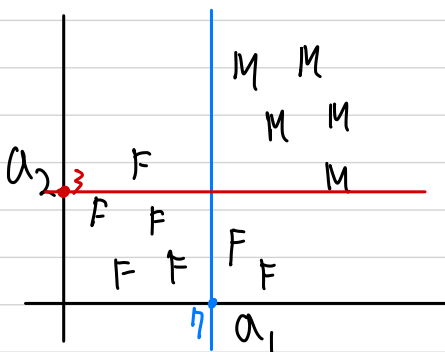
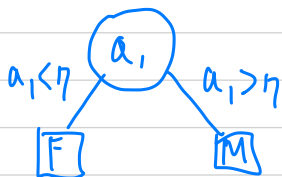
linear model



-Decision Tree-



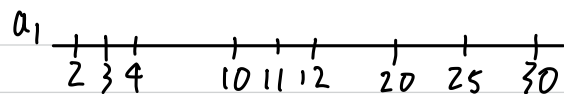
엔트로피를 사용하여 a_1 , a_2 둘중 무엇을 먼저 사용하여 나눌지를 결정한다.



-군집화 : K-means 알고리즘-

a_1	a_2	a_3	a_4	a_5	class
2					
4					
10					
12					
3					
20					
⋮					

로이드가 만들었다. 상당히 간단하고 상당히 강력하다.



클러스터가 몇개가 만들어질 수 있을까?

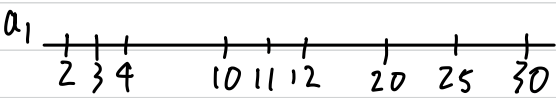
몇개의 클러스터가 적절할까?

$k=2$ 라면?

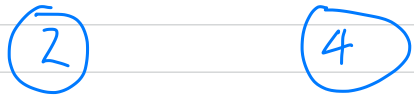
이 알고리즘은, 학습방법을 사용,

처음에는 랜덤으로 클러스터를 시작하는데,

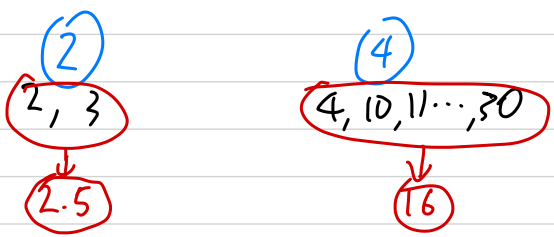
점차 반복(iterative하게)하면서 효율적인 클러스터를 (학습하여) 만들어 나간다.



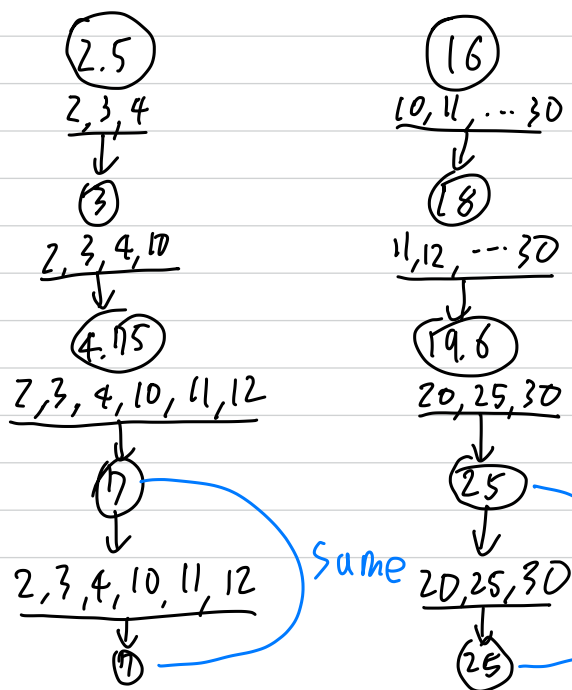
1. random으로 평균값을 2개를 설정한다.



2. 그 mean에 가까운 숫자를 가져온다.



3. 다시 평균을 계산한다.



Same → Same 한 means이 나오면 중단.

2차원은 어떠한가?



어느것이 중심으로 부터 가까운지 여부가 cluster의 여부가 결정된다.

이 알고리즘의 문제는 random한 초기 mean값에 의해

cluster의 성능, 결과가 달라진다는 것이다

- 연관분석 A-priori 알고리즘 -

이 알고리즘은 경영에서의 장바구니 분석에서 시작되었다.

예를 들어, A와 F가 같이 잘 팔리더라.

그러면 A와 F를 같이 진열하던가, 같이 사면 세일을 해준다던가 등등

	물건1	물건2	물건3	...	물건30
t1	1	0	0	...	1
t2	0	1	1	...	0
⋮	⋮	⋮	⋮	⋮	⋮
tn	0	0	1	...	1

여러번 같이 발생한 것을 찾고 싶은 것이다.

Transaction Item -> Frequent Item set -> Association Rules

자주 발생한 빈도

minimum support 이상 발생하면 '자주'라고 표현

ex) 2000만번 이상 (카운트)
20%

Transaction Items \longrightarrow Frequent Item sets

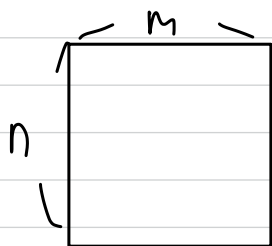
1 {1,2,3}	minimum support : 2	{1, 3}
2 {1,4}		
3 {1,3}		
4 {2,5,6}		

\longrightarrow Association Rules

If 1, then 3 : confidence 66%

If 3, then 1 : confidence 100%

문제는 알고리즘 관점에서 complexity가 너무 크다. : $O(2^n)$



$m C_2$ 번 minimum support가 넘는지를 봐야한다.

그걸 또 n 번 계산해야 한다.

$$= O\left(n \cdot \frac{m(m-1)}{2}\right)$$

$$\sum_{i=1}^m m C_i \cdot n = \underline{\underline{2^{m-1} \times n}}$$

왜 이 문제가 어려운가?

장바구니들에서 이런 아이템이 같이 팔려요!

= 문서들에서 같이 발생하는 단어가 있어요!

와 같은 의미이다.

같이 발생한다는 것을 알아내면?

: 검색 알고리즘에서 사람들이 연관된 단어들을 검색창에 즉시즉시 띄워주는 것이다.

같이 물려다니는 웹페이지들을 생각해보자.

: 같이 연관된 페이지들을 추천해주는 것이다.