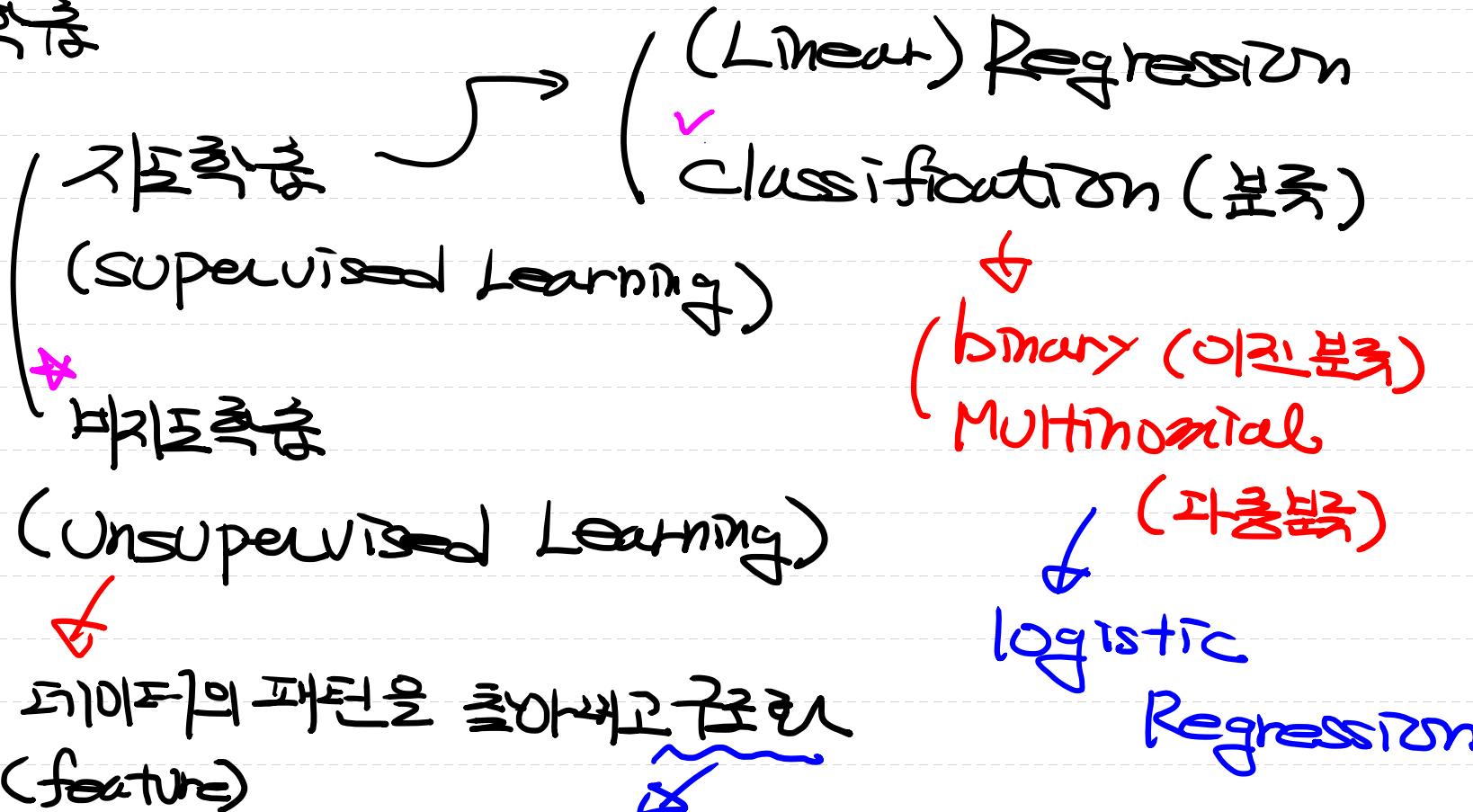


• 12/11

## • 비지도학습



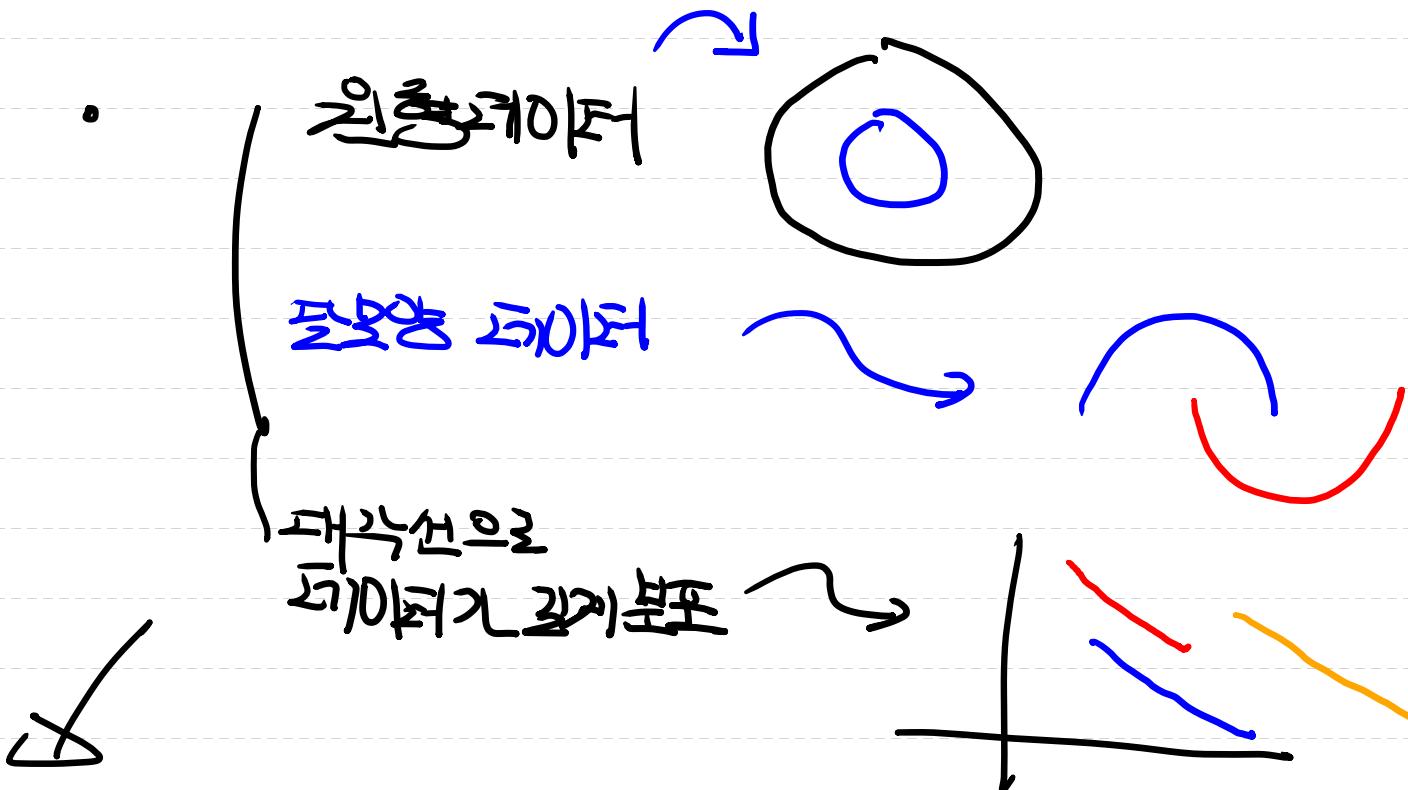
비지도학습

→ 알고리즘

- K-means
- DBSCAN

- k-means

- ① k 값을 hyperparameter.
- ② 구간의 중앙  $\rightarrow$  centroids
- ③ EM 알고리즘을 기반으로 동작!
- ④ 처음 random하게 초기화하는 centroids가  
파괴 결과가 솔이해요  
 $\rightarrow$  여러번 알고리즘을 수행해가며 가장 좋은  
결과를 사용 (Majority Voting)
- ⑤ code로 알아보아요~ (Iris)
- ⑥ 한계  $\rightarrow$  데이터의 분포가 특수한 형태일  
경우 “2각기반” 이기 때문에 잘  
동작하지 X



한계를 해결하기 위해

‘DBSCAN’ → density(밀도)를 기반으로 clustering!

hyper parameter  
 $\epsilon$  (epsilon: cluster의 브경)  
 $\text{minPts}$ : cluster 안의 개체의 최소주

↳ k-means 알고리즘으로 해결하지 못한  
 해결 가능

## Dimension Reduction (차원축소)

↳ 차지도 학습할 때 더 많이 사용되요!

고차원 데이터 → 저차원 데이터

\* ↓  
풀연산으로 정보丟失 (Information loss)

이거 왜 해야 돼요??

고차원 데이터는 데이터가 희박해요 (Sparse)

↳ 데이터 불균형, 과적학습이 잘 되지 않아요

\* 속도가 좋아요

, Overfitting을 줄일 수 있어요

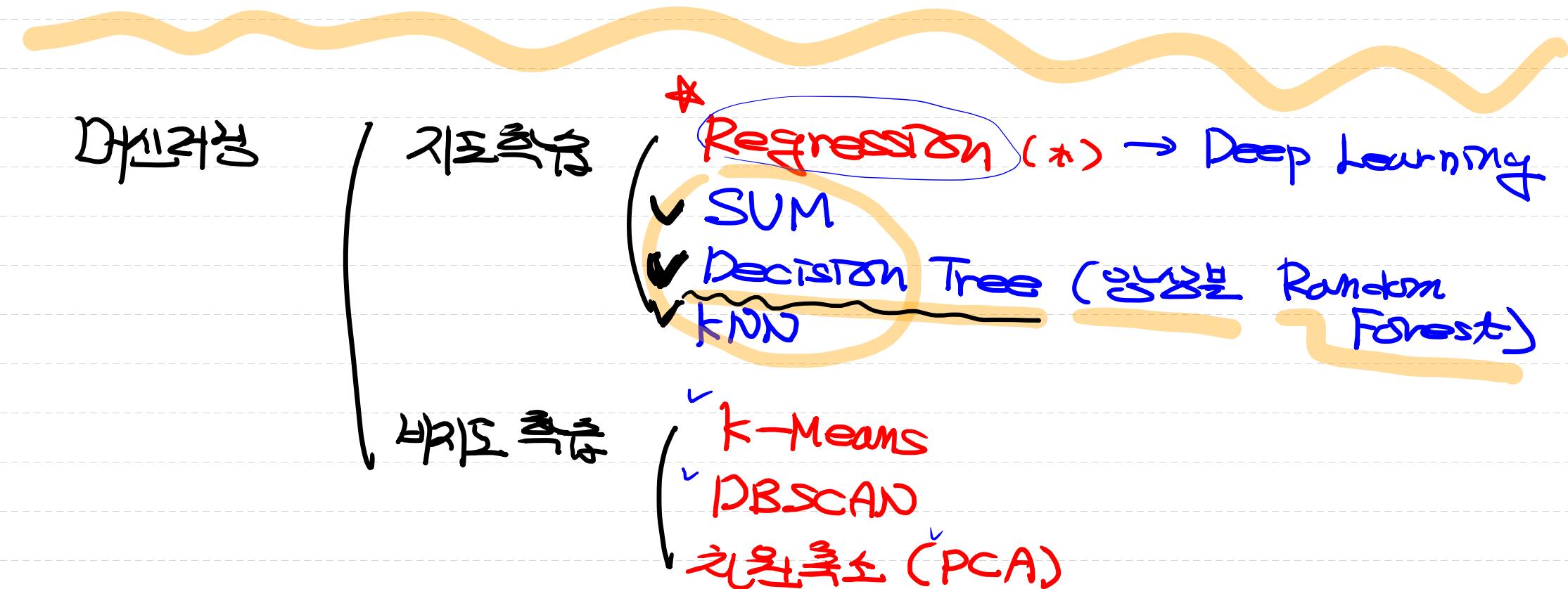
\* 연산의 속도가 높아져요!

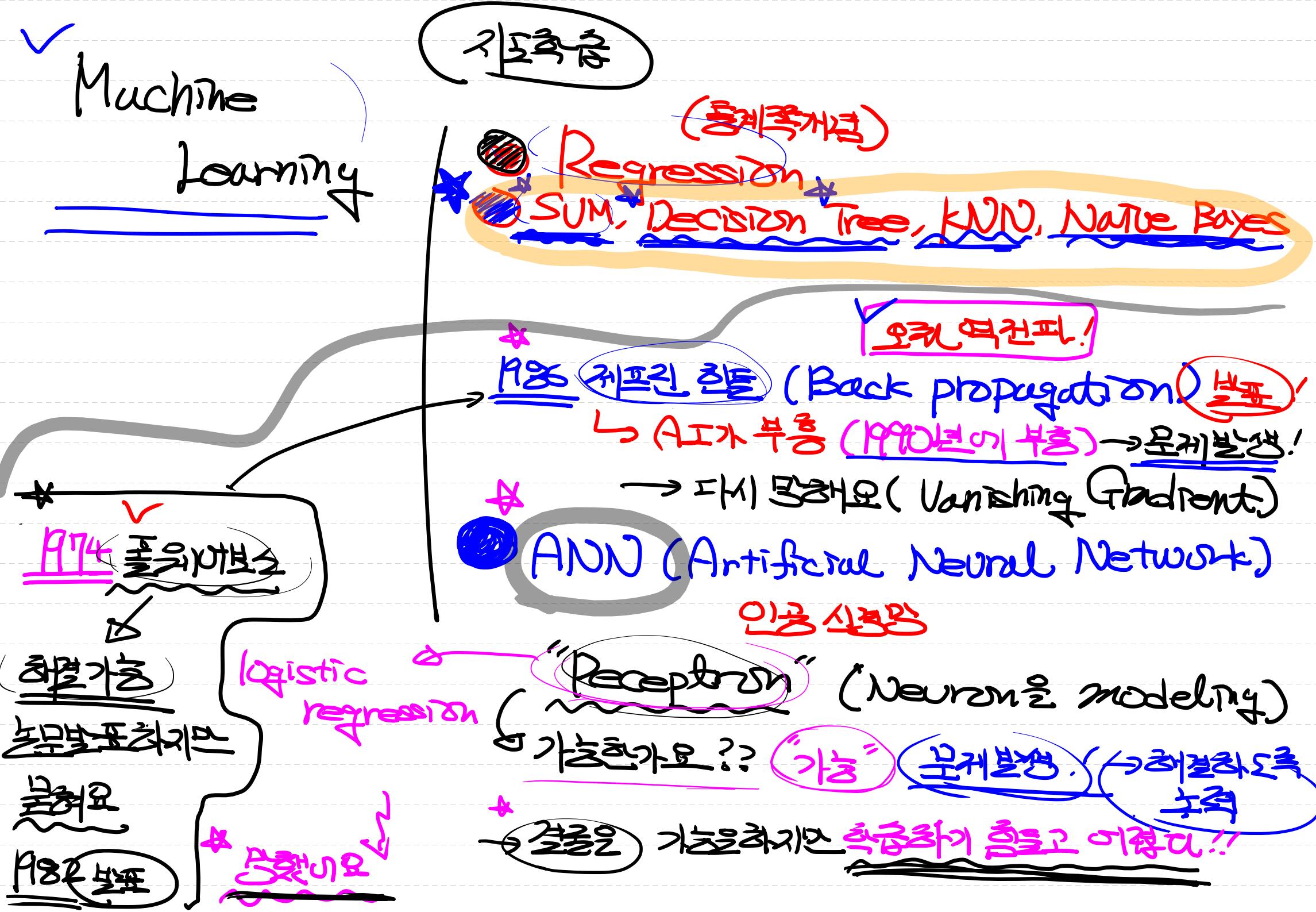
## ④ 차원축소 알고리즘

↳ PCA (Principle Components Analysis)  
(! 주성분분석!)

설계로 사용하기 어렵거나 짜증나면 X → library를 사용

(sklearn)





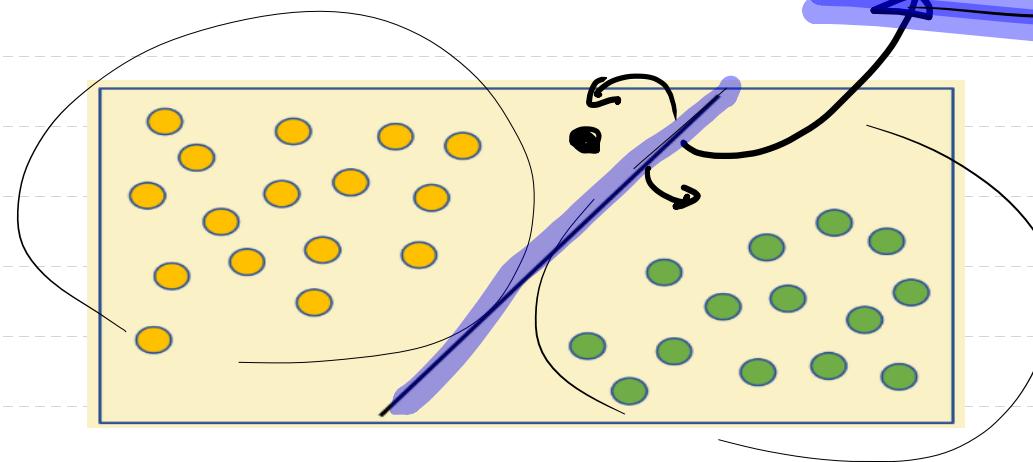
● 머신러닝 알고리즘 → ① SVM (Support Vector Machine)

(regression 문제)  
classification 문제

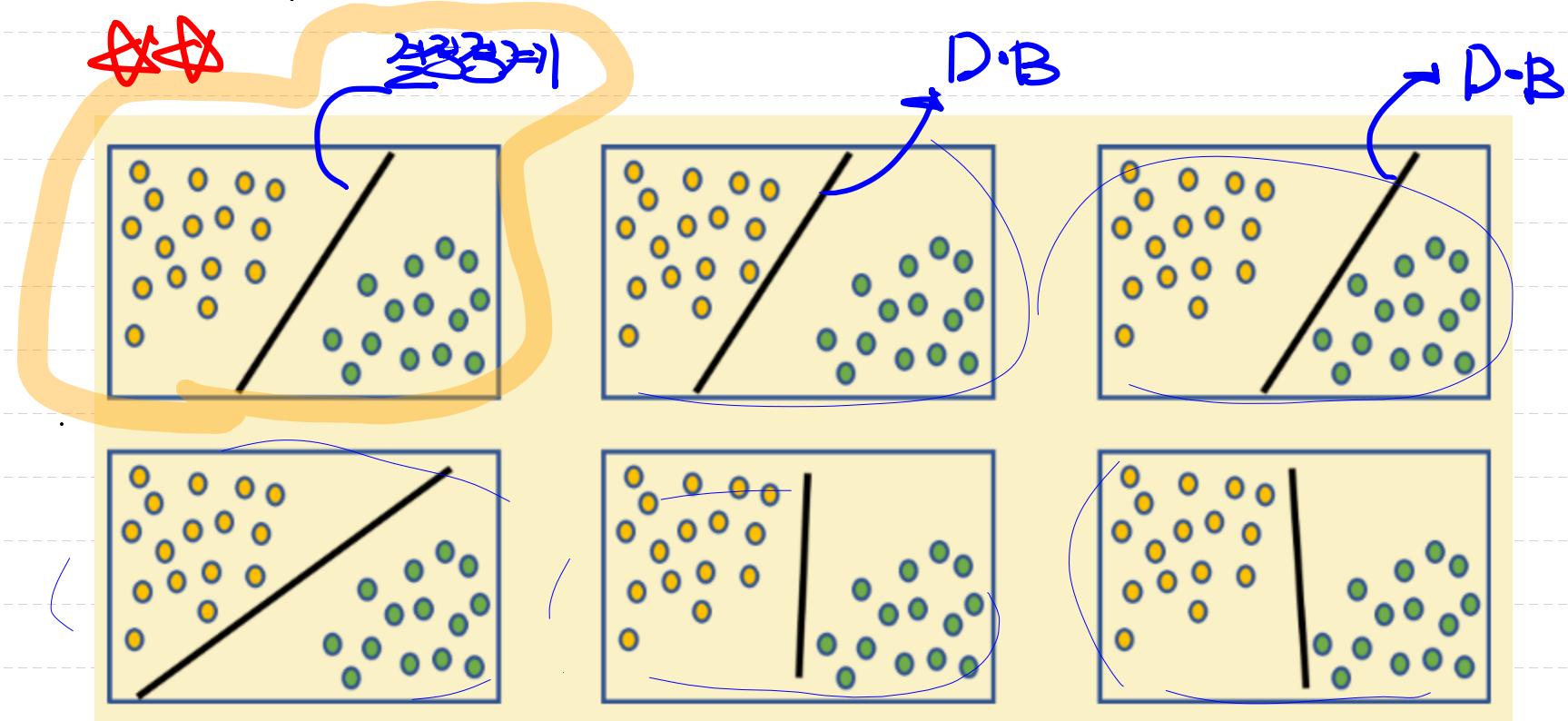
둘 다 사용 가능

성능은 DNN보다 훨씬 높지만 model이 가볍고 빠르죠!

SVM → Decision Boundaries (결정 경계)를 찾는  
방법이 있나요?



## Decision Boundaries (결정경계)



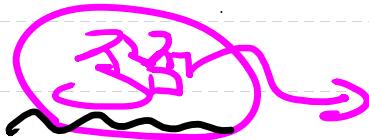
여러가지 결정경계 (Decision Boundaries) 중 어떤게 가장

좋나요 ??

\* Support Vector Machine (SVM)

이거 몰라요 헝

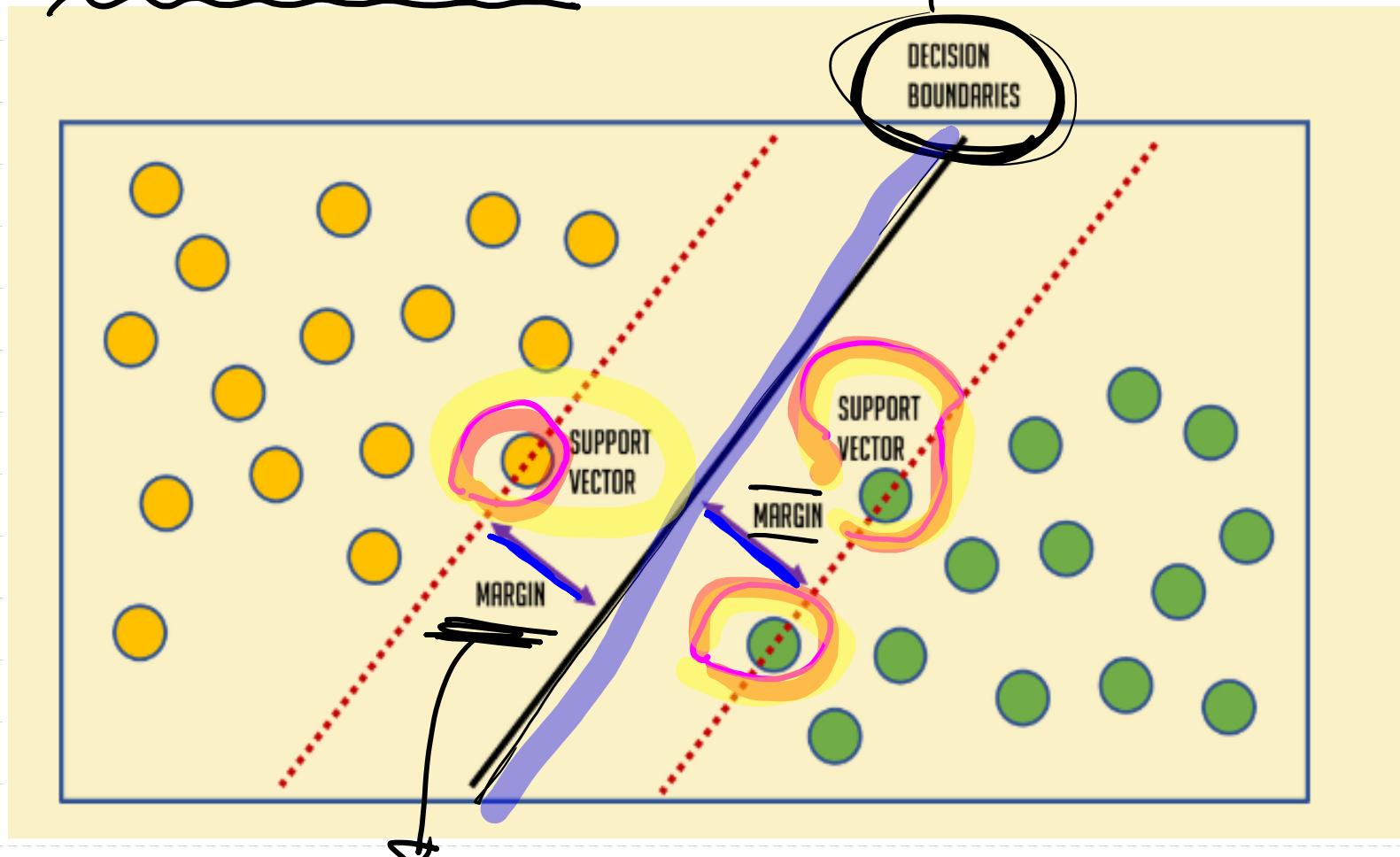
SUM ↗



수직으로 속도가 빨라요!

↗

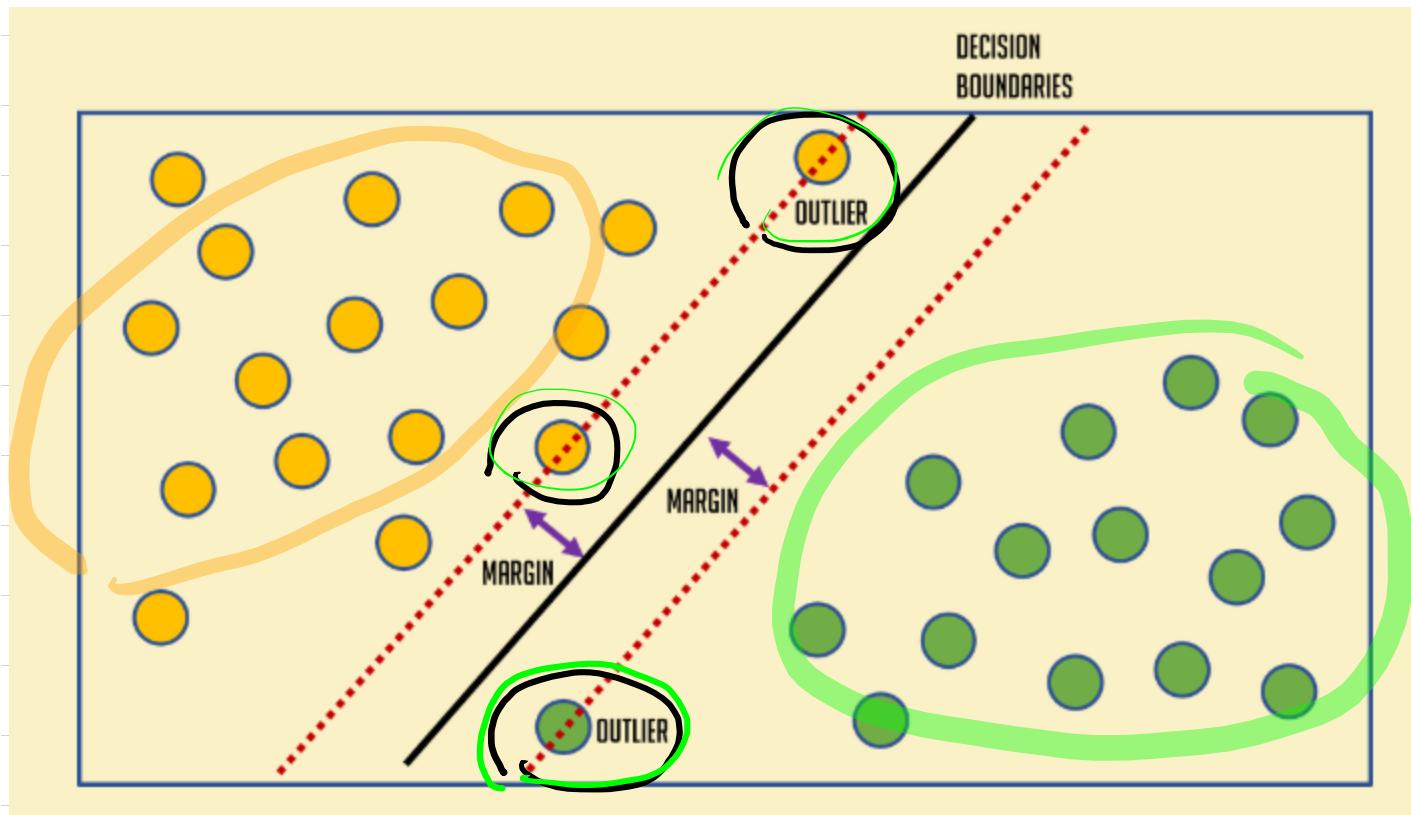
이전을 훑는게 목표!



Support Vectors와 Decision Boundaries의

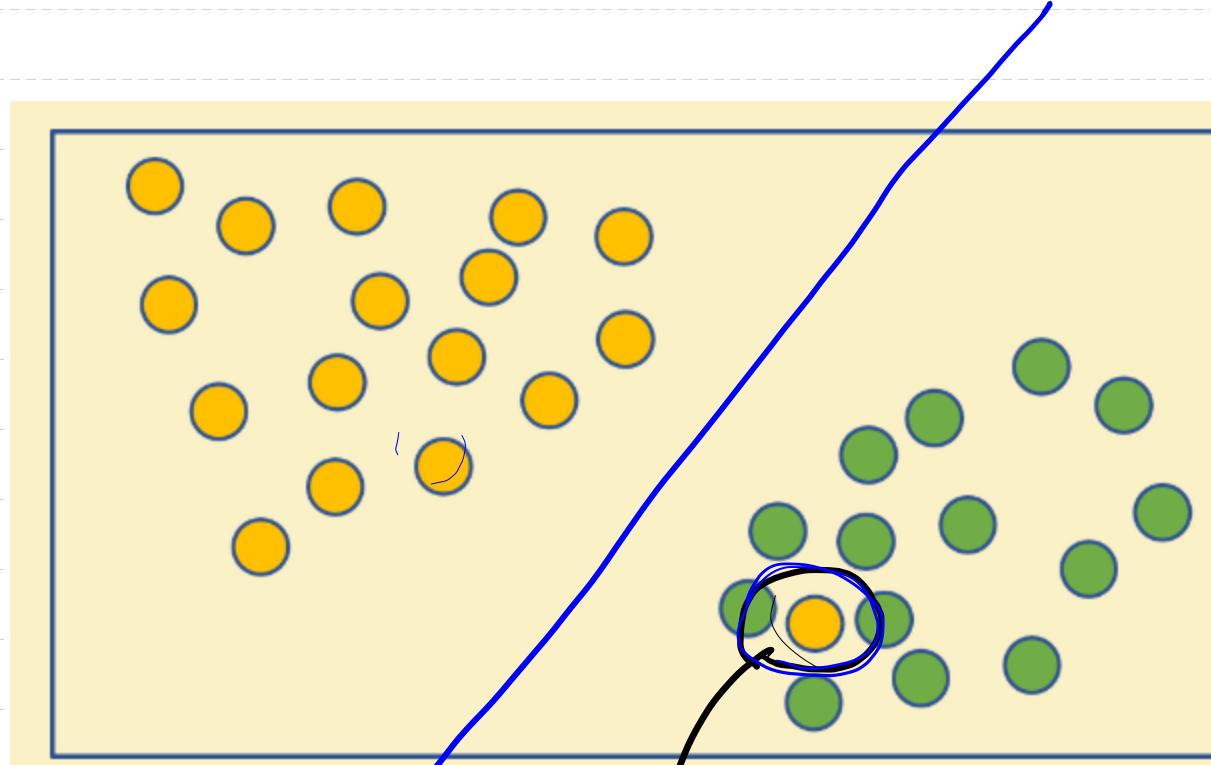
거리 → 이거 최대값 되는 거  
우리가 원하는 경계!!!

- SUM이 n이 짐작하지 못하는 경우! → 이상치가 아주 많음.



이상치가 존재하면 Support Vectors가 이상치로 결정!

## ◎ 조금 더 심한 경우!

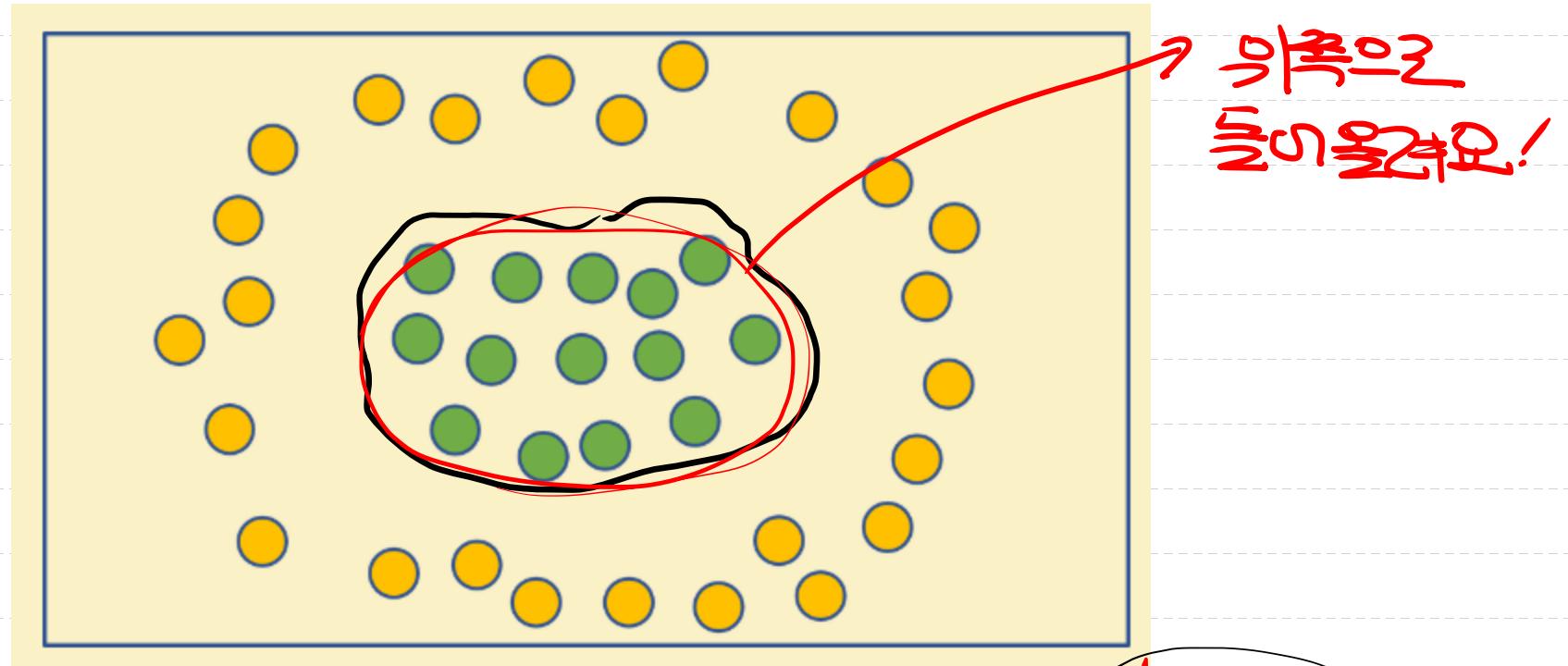


이전경우와는 “정용규칙”  
(Regularization)  
↳ 얼마나 정용화 되는가?

정용화

이걸 이용하면  
가능

그럼 이런 경우에는 어떻게 해나요?



이런경우에는 kernel 기법을 이용.

(linear → 직선  
poly  
rbf → 고차원으로)

이런경우에는 Data를 차원이 고차원으로

Projection 시켜요

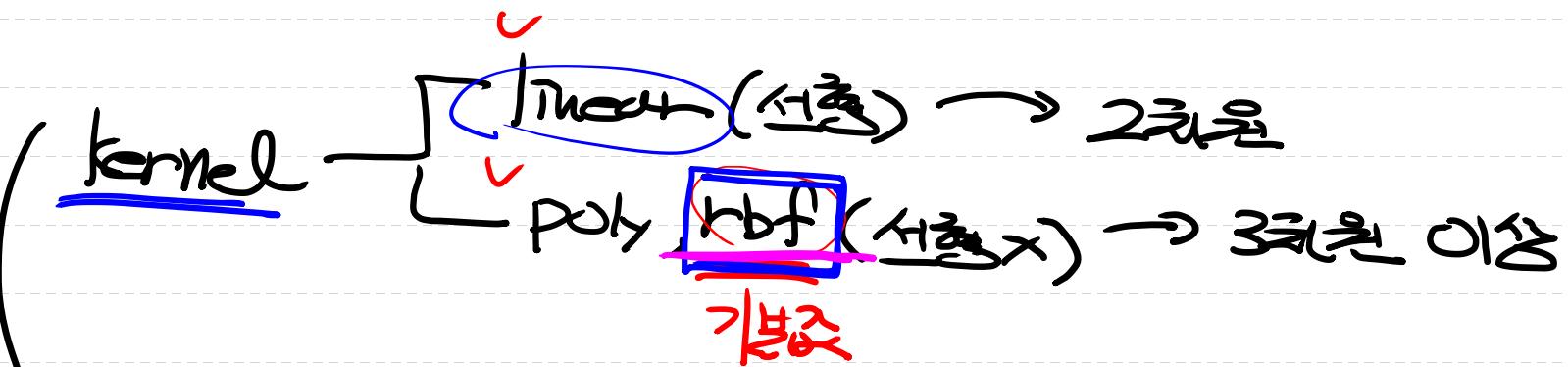


SUM

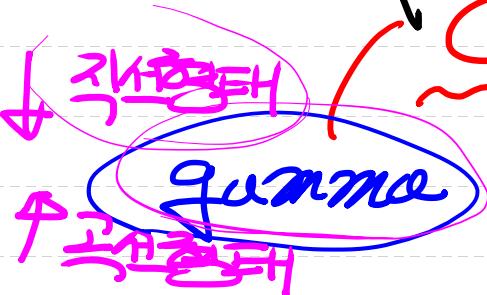
↳ Sklearn을 이용하면 간단하게 M-L 구조이 가능

Parameter를 잘 설정해야해요!

SUM



Cost 흐름 (기본값 1)



Cost의 값이 클수록 다른 범주에 놓이는 data 포인트를  
적게 해요  
→ Overfitting의 위험.

SUM을 이용하여 Cost값  
구함(rbf가 블랙박스 등)

이 hyperparameter를 잘 조정해야 해요.

→ 이 값을 바탕으로 우리 model을 훈련해야 하니까

수작업으로 하면 너무 힘들어요!!

↳ 자동화하면 어떻게 해야 하나요?



"  
코스로 알아보아요!!" → "Grid SearchCV"