

SVM(Support Vector Machine)

Classification/Regression 둘 다 사용될 수 있지만 주로 Classification에 사용된다고 한다.

모든 피처를 사용!

Supervised Machine Learning Algorithm!

선형/비선형
Binary/Multinomial

YBIGTA 디자인팀 12기 김현주

마진

◇ 마진: 초평면과 각 클래스 별 초평면 사이 거리가 가장 가까운 데이터

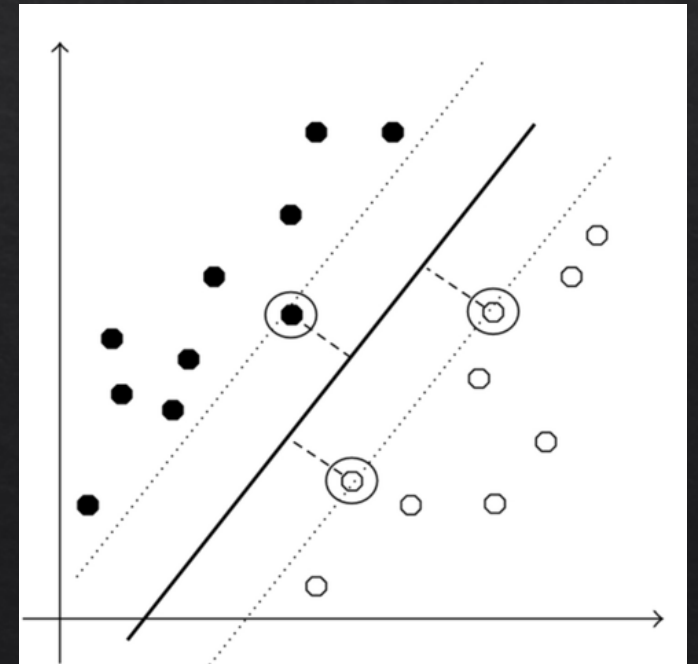
◇ SVM은 이 마진을 가장 크게 만드는 선을 찾는 것을 목표로 한다.

→ 최대마진초평면 = 최적의 초평면

* 초평면: N차원에서 N-1차원의 공간!

(N차원을 두 공간으로 나누게 된다.)

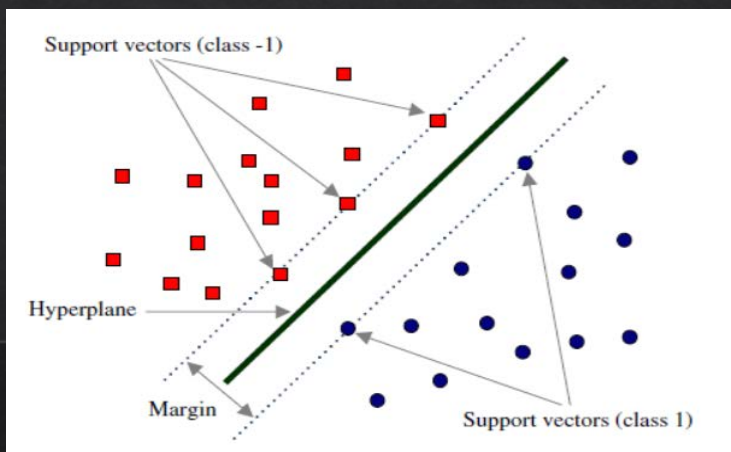
* 그림에서 표시된 점들과 초평면 사이의 거리가 마진!



하드 마진, 소프트 마진

하드 마진

- ◆ 데이터가 오류없이 하나의 선형모형으로 구분될 수 있을 때 사용
- ◆ 수많은 초평면 후보 중 가장 큰 마진을 가지는 초평면을 고르자. 가장 안정적으로 완벽하게 구분할 수 있는 초평면! → 최대마진초평면



소프트 마진

- ◆ 데이터가 선형모형으로 구분될 수 없을 때 사용
- ◆ 핵심은 여유변수!(선형으로 분리가 불가능할 때 오류가 생길 수 밖에 없는데, 어느 정도의 오류를 허용할 것인가!) → 제약식을 추가하자

$$\text{Objective Function} = \min_{w, b, \xi} \frac{1}{2} \|W\|^2 + C \sum \xi_i$$

So that,

$$y_i(w_{i1}x_{i1} + w_{i2}x_{i2} + b) \leq 1 - \xi_i \quad \text{for all } y_i$$

- ◆ C가 커지면 제약을 많이 해서 마진 거리가 줄어들게 된다.(오류 범위 좁아짐, 보수적)
- ◆ C가 작아지면 제약을 적게 해서 마진 거리가 늘어나 오류 범위를 넓혀준다.

비선형 SVM

- ◆ 비선형 SVM: Input Space에 있는 데이터를 고차원 공간(Feature Space) → 매핑(Mapping) → 두 범주로 분류하는 선형 초평면
- ◆ 매핑함수: 2차원 → 3차원 함수 ϕ .
- ◆ 선형결정면 정의식:
$$f(x) = \sum_{i=1}^n \alpha_i \phi(x_i) \phi(x) + b$$
- ◆ 위 정의식의 두 파이 함수(커널함수)의 내적으로 구할 수 있다.

