

3.4~3.7

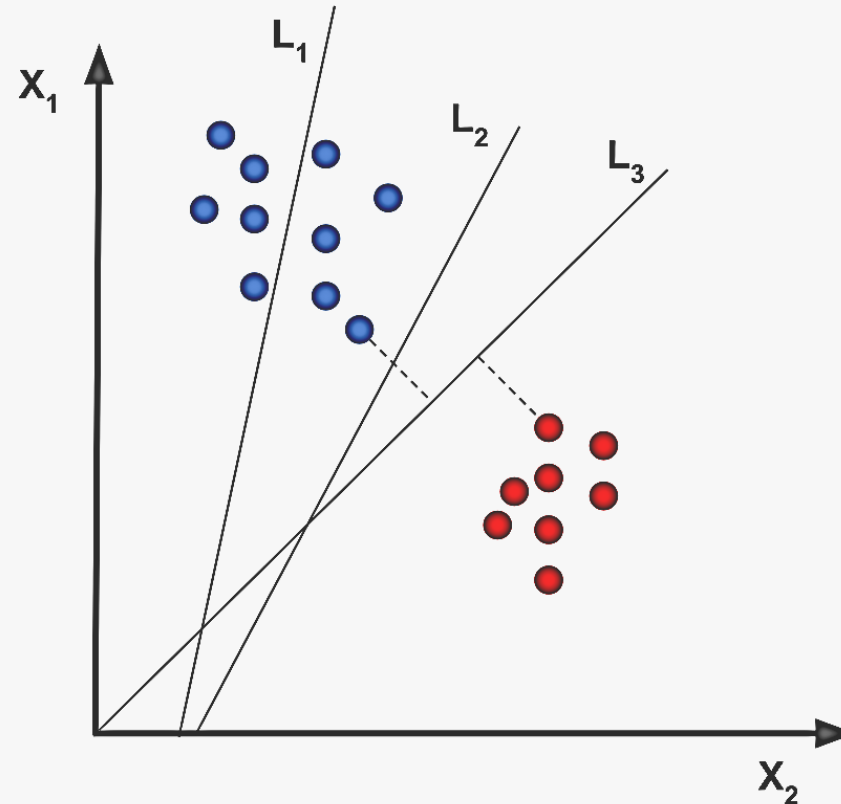
Chapter 3. 분류

Contents

- SVM – 결정경계, 마진
- 선형분류기
- OvO VS OvA

Support Vector Machine

IDEA : "How do we **divide** the space with decision boundaries?"



Support Vector Machine

- 기존의 데이터로 분류 기준을 정함
- 새 데이터가 들어왔을 때 어느 그룹에 속하는지 **예측**
- 비확률적 이진 선형 분류 (왜 비확률적인가?)
- 분류 기준으로 사용되는 것 : 결정 경계 (기준 선 역할)

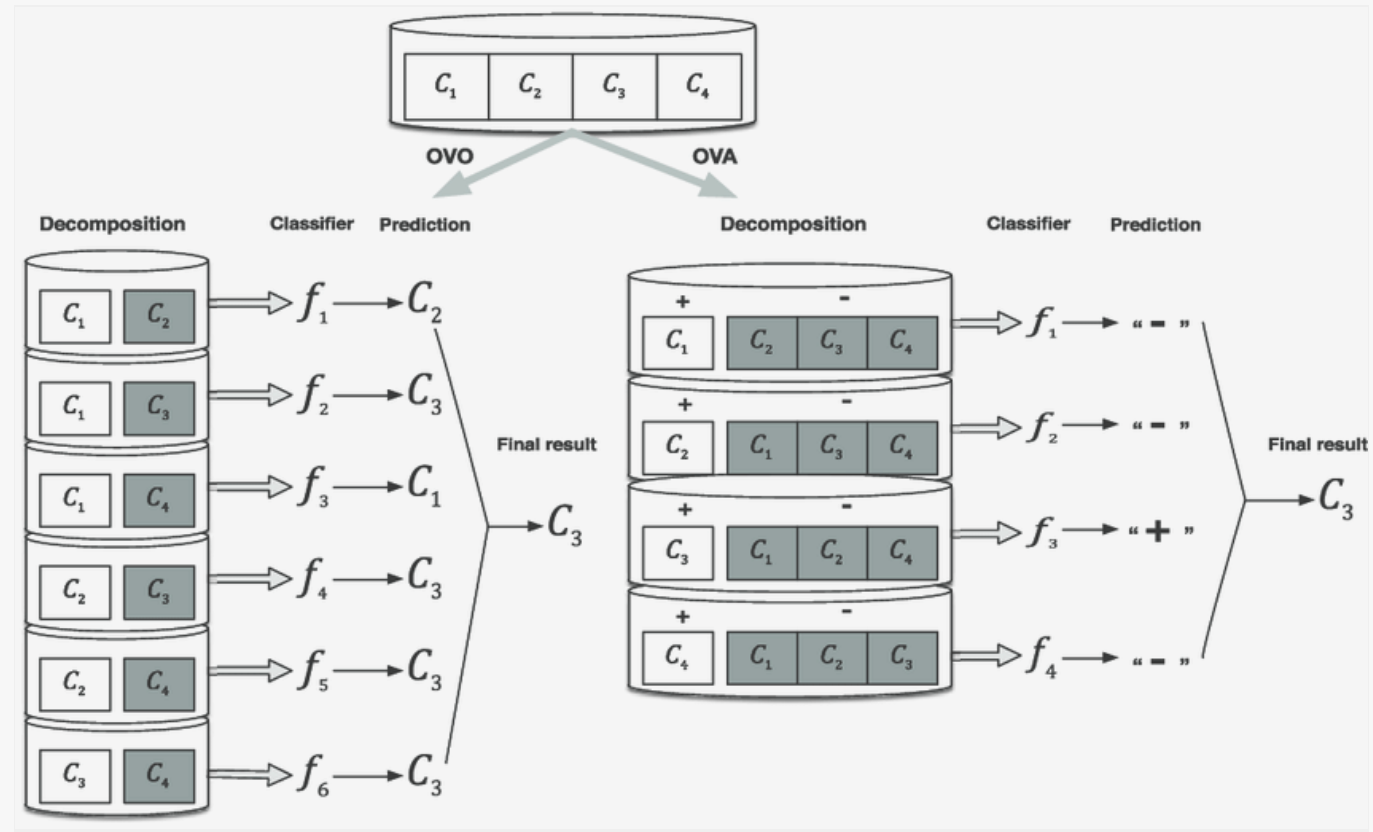
결정 경계(Decision Boundary)

- 마진이 가장 큰 결정 경계를 찾는다
- 고려 사항 :
 - 1) 이상치 조정을 통한 결정경계 정의 :
 - Hard Margin : SVM의 이상치 허용 X
 - Soft Margin : SVM의 이상치 허용
 - 2) 차원에 따른 결정경계 정의 : Hyperplane (초평면)

Hard Margin VS Soft Margin

- Hard Margin :
 - 엄격히 클래스를 분리 → 모든 입력 tuple은 무조건 한 클래스에 속함
 - 노이즈로 인해 결정경계를 찾기 어려울 수 있음 → 현실적으로 사용 어려움
- Soft Margin :
 - 경계선에 약간의 여유 변수(Slack Variable)를 둠
 - 얼마만큼의 여유를 가지고 오류를 인정할 것인가?

OvO VS OvA



OvO (One versus One)

- 모든 가능한 두 개의 클래스 조합에 대해 이진 분류기를 만듦. 가장 많이 양성으로 분류된 클래스 선택
- SVM의 경우 기본적으로 LinearSVC(liblinear library 사용)를 제외하곤 OvO 방식으로 작동합니다

OvA (One versus All)

- One VS All 비교하는 classifier를 각각의 class마다 학습
- 모든 classifier 중 가장 큰 예측값을 가지는 class를 최종값으로 선택

^~^