

SVM support Vector machine

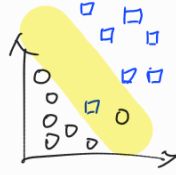
- 다양한 머신러닝 알고리즘 \Rightarrow 선형/비선형, 분류/회귀, 이상치 탐색 모두 사용 가능.
- 복잡한 분류 문제에 유용
- 크고 많은 데이터셋에 적합. (클래스 크기 데이터)

기본 선형 모델 $\begin{cases} \text{경사하강법} - \text{회귀} \\ \text{로지스틱 회귀} - \text{분류} \end{cases}$

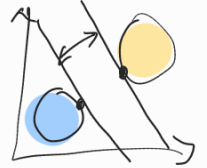
선형 SVM

- 분류

- class 사이 가장 넓은 공간 찾기
마진 분류



양쪽 경계에 위치한 샘플로 결정
 \Rightarrow support vector.



- 소프트 마진 분류

모든 샘플이 선형 데이터로 잘 나열되어 있는 경우 하드 마진 분류

\Rightarrow 하드 마진인 경우 굳이 머신러닝 안해도 알 수 있을 정도.

① / ②
53.

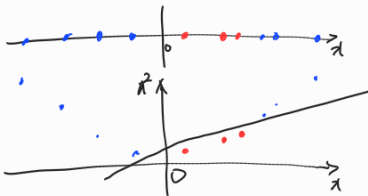
선형적, 이상치 민감.

(① 도록 최대한 넓게
② 마진 오류 사의 적절한 균형) \Rightarrow 소프트 마진 분류

비선형 SVM

선형 $ax+c$
비선형 $ax^n + \dots + c \quad (n \geq 2)$

- 다항 특징 추가



알고리즘 그대로 두고
데이터를 변형
(데이터의 차수 높임)

높은 차수 : 데이터 잘 표현되리 못함.
높은 차수 : 계산이 복잡 (속도 \downarrow)

- 커널 트릭 (kernel trick)

- 다항 특징 추가 시 계산량 폭증 문제 (시간 \uparrow) \Rightarrow 계산량 줄이는 방법?

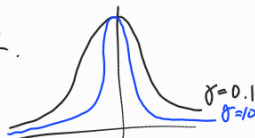
- 변환 함수 존재한다. 다항 특징 추가하지 않아도 다항 특징 여러 개 추가한 것과 같은 결과

- 다항식 커널 : 다항식을 추가한 것과 같은 효과내는 커널

- 유사도 특성

- 샘플이 렌더마크 (특정 모양)과 얼마나 닮았는지 측정하는 유사도 함수로 계산한 특성 추가

- 가우시안 RBF 정규 분포.



SVM 회귀

- 제한된 마진 오류 안에 가능한 많은 sample 담기

