

Pair Programing

6주차

1조
201984014
성도윤
201904126
허준혁

Q1. 하드 마진 분류

A1. 데이터가 선형적으로 구분되는 경우에만 작동 된다
-> 이상치에 민감해서 잘 안쓰일거 같다

Q2. 소프트 마진 분류

A2. 도로폭이 넓어지면 넓은 마진 오류, 도로폭이 좁아지면 적은 마진 오류
-> 넓은 도로에 적은 마진 오류가 좋다
-> 적절한 값을 찾는 것이 중요하다

Q3. 유사도 특성

A3.
개념 : 데이터 포인트 간의 유사성을 나타내는 측정 값

활용 분야 : 텍스트 분석에서 문서 유사성을 측정하여 문서 분류, 검색 및 군집화 작업을 수행하는데 사용

장점 : 차원이 커지면서 선형적으로 구분될 가능성이 높음
단점 : 훈련 세트가 클 경우 동일한 크기의 아주 많은 특성 생성

Q4. 다항식 커널

A4. 비선형 관계를 처리하기 위한 강력한 도구

-> 계산 비용 발생

-> 하이퍼 매개변수 조정 및 정규화 필요

Q5. 머신러닝에서 커널이란

A5. 변환을 계산하지 않고 원래 벡터에 기반한 점곱을 계산할 수 있는 함수

Q6. SVM이란?

A6. 분류 및 회귀 작업에 사용되는 지도형 기계 학습 알고리즘의 유형

-> 구성 요소에는 어떤 것들이 있을까?

-> 선형 분리, 마진, 커널 트릭, C매개변수, 소프트 마진 등이 있다

-> SVM을 쓰는 이유는?

-> 다양성과 효율성이 좋다

Q7. SVM 분류와 회귀

A7.

SVM 분류 : 마진 오류 발생 정도를 조절하면서 두 클래스 사이의 도로폭을 최대한 넓게 하는 것

SVM 회귀 : 제한된 마진 오류 안에서 도로 안에 가능한 한 많은 샘플 포함하는 것

-> 회귀를 처리 할 때는 커널 SVM 모델을 사용

Q8. 온라인 학습이란?

A8. 새로운 샘플에 대해 점진적으로 학습하는 것

-> 지속적인 샘플 데이터 필요한가?

-> 온라인으로 데이터를 가져오는 방식 같다

-> 대규모의 비선형 문제라면 신경망 알고리즘 고려