SVM 요약

2025170941 이준서

서포트벡터머신:

지도학습의 한 종류로 주로 이진 분류 문제를 해결하는 데 사용

그러나 다중 분류와 회귀문제에도 확장 가능

Separating Hyperplane:

클래스 간의 경계를 가장 넓게, 즉 마진을 최대화하는 초평면(hyperplane)을 찾는 것. 예를 들어 데이터가 두 클래스로 나눠져 있다고 할 때, 이 사이에는 여러 개의 경계선이 존재하고 그중에서도 양쪽 클래스와의 거리를 가장 멀게 유지하는 초평면을 정해야 한다. 그래야 일반화의 오류가 최소화되며 좋은 예측이 가능해진다.

여기서 마진은 각 클래스에서 가장 가까운 관측치 사이의 거리를 나타내며 w(기울기)로 표현된다. w는 행렬로 표현되어 아직 내가 알지 못하는 수학으로 수식 전개는 넘어가겠다.

Support Vectors: 결정 경계 근처에 존재하는 가장 중요한 데이터 포인트들을 서포트 벡터라 부른다.

Linearly Nonseparable Case:

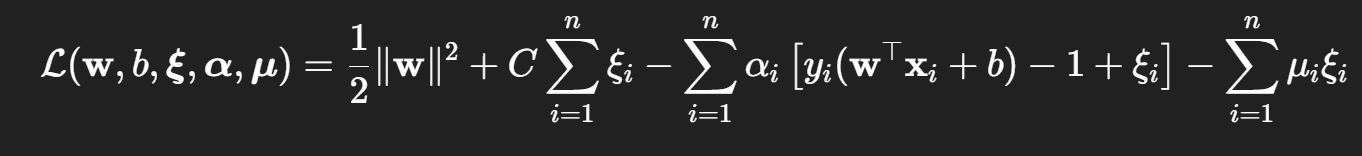
선형적으로 완벽히 구분되지 않는 경우가 존재한다. 이때 Soft Margin SVM을 사용한다. 완벽한 선형 분리를 강제하지 않고 일부 오분류를 허용하되 마진을 최대화하는 방식으로 오차를 허용한다는 특징이 있다.텍스트, 폰트, 그래픽, 타이포그래피이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.다음과 같은 목적 함수를 사용한다. ||w||^2은 마진의 역수, 시그마로 나타난 합은 마진을 위반한 정도의 총합이며 C는 규제 파라미터로 마진과 오차 간 균형을 조절한다.

텍스트, 폰트, 그래픽, 블랙이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.다음은 제약조건이다. 각 데이터 포인트가 마진에서 얼마나 벗어나는지 조절한다. ξi = 0 이면 완벽히 마진 바깥쪽에 분류된 것, 0< ξi <1 이면 마진 안쪽에 분류, ξi > 1이면 잘못 분류된 것이다.

Lagrangian Formulation: 라그랑지언 정식화



제약 조건을 라그랑주 승수와 함께 붙인 것이다. 라그랑주 승수란 제약 조건이 존재하는 최적화 문제를 풀 때 사용하는 변수다. 지금 여기선 벡터로 표현되었다. 알파는 클래스 분리 조건을, u는 슬랙 변수>=0 임을 제약한다.

KKT 조건:

최적화 문제를 풀기 위해 필요한 조건들이다.

텍스트, 스크린샷, 폰트, 도표이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.

Kernel Methods for Nonlinear Classification:

커널 메소드를 활용한 비선형 분류기는 클래스가 선형적으로 분류되기 어려울 때 쓰인다. 예를 들어 원 안팎으로 클래스가 나뉘는 데이터가 있다고 하면, 선형 초평면으로 분류하기 힘들기 때문에 커널 메소드를 활용한다. 선형으로 분리되지 않는 데이터를 고차원 공간으로 옮겨서 선형 분리를 꾀하는 방법으로 입력 벡터를 고차원 공간으로 보내는 매핑 함수가 쓰인다.