5장 서포트 벡터 머신 (2부)

주요 내용

- SVM 이론

5.4 SVM 이론

결정 함수와 예측

- 결정 함수: 아래 값을 이용하여 클래스 분류

$$h(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b = w_1x_1 + \cdots + w_nx_n + b$$

- 예측: 결정 함수의 값이 양수이면 양성, 음수이면 음성으로 분류

$$\hat{y} = \begin{cases} 0 & \text{if } h(\mathbf{x}) < 0 \\ 1 & \text{if } h(\mathbf{x}) \geq 0 \end{cases}$$

결정 경계

- 결정 경계: 결정 함수의 값이 0인 점들의 집합

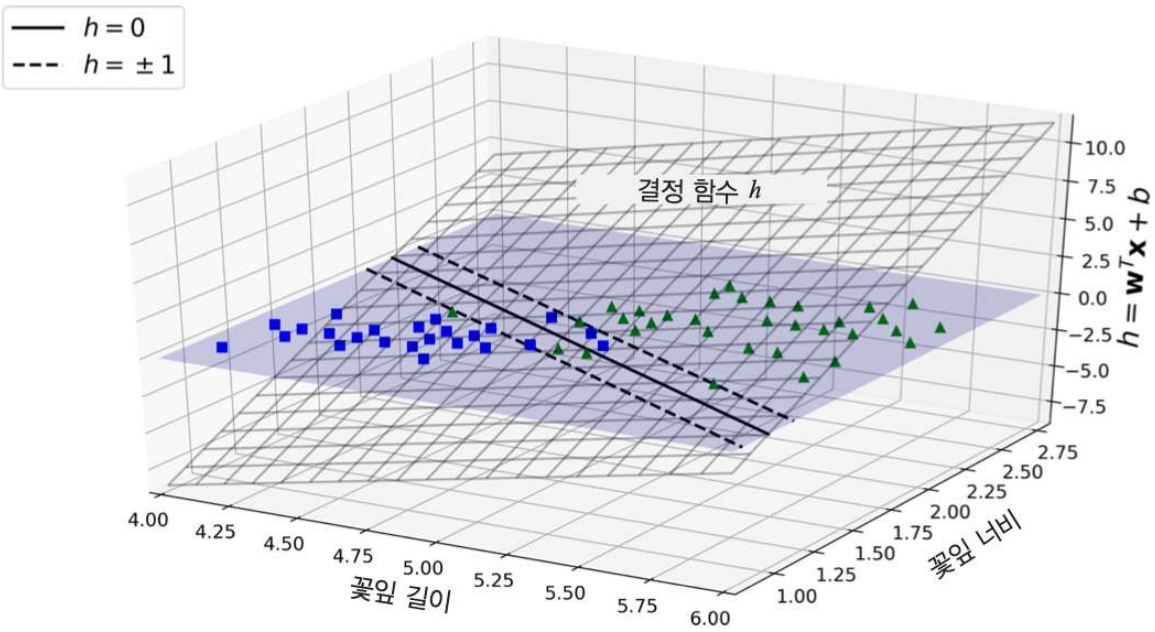
$$\{\mathbf{x} \mid h(\mathbf{x}) = 0\}$$

- 결정 경계 도로의 경계: 결정 함수의 값이 1 또는 -1인 샘플들의 집합

$$\{\mathbf{x} \mid h(\mathbf{x}) = \pm 1\}$$

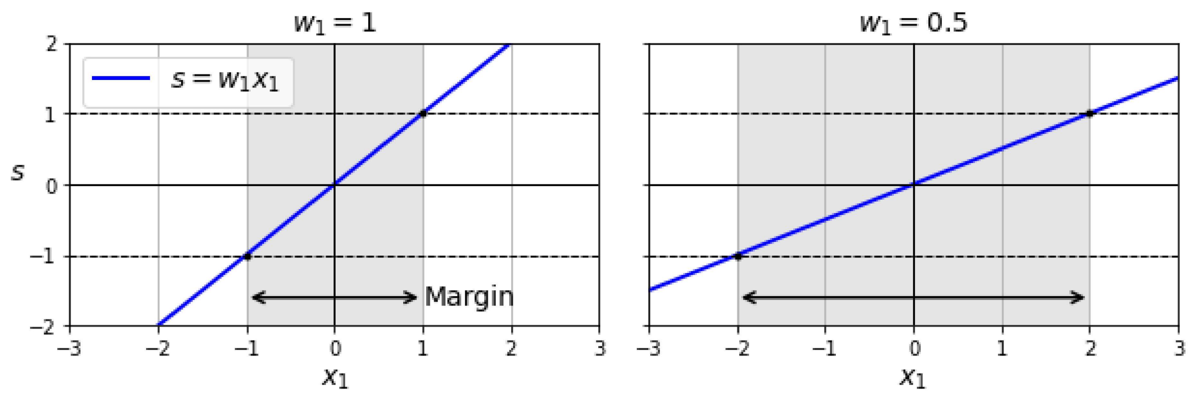
예제

붓꽃 분류. 꽃잎 길이와 너비를 기준으로 버지니카(Iris-Virginica, 초록 삼각형) 품종 여부 판단



결정 함수의 기울기

- 결정 경계면(결정 함수의 그래프, 하이퍼플레인)의 기울기가 작아질 수록 도로 경계 폭이 커짐.
- 결정 경계면 기울기가 $\|\mathbf{w}\|$에 비례함. 따라서 결정 경계 도로의 폭을 크게 하기 위해 $\|\mathbf{w}\|$를 최소화해야 함.



- 하드 마진 모델 훈련: 모든 양성(음성) 샘플이 결정 경계 도로 밖에 위치하도록 하는 기울기 찾기.
- 소프트 마진 모델 훈련: 결정 경계 도로 위에 위치하는 샘플의 수를 제한하면서 결정 경계 도로의 폭이 최대가 되도록 하는 기울기 찾기.

목적 함수

- 결정 경계면의 기울기 $\|\mathbf{w}\|$를 최소화하는 것과 아래 식을 최소화하는 것이 동일한 의미임. 따라서 아래 식을 목적함수로 지정함.

$$\frac{1}{2}\|\mathbf{w}\|^2 = \frac{1}{2}\mathbf{w}^T\mathbf{w}$$

- 이유: 함수의 미분가능성 때문에 수학적으로 다루기가 보다 쉬움. $1/2$ 또한 계산의 편의를 위해 추가됨.

하드 마진 선형 SVM 분류기의 목표

- 목적함수를 최소화하는 파라미터 벡터 $\mathbf{w}$를 구하기 위해 다음 **최적화 문제**를 해결해야 함. 즉, 모든 샘플 $\mathbf{x}^{(i)}$에 대해 만족시켜야 하는 조건이 추가되었음.

$$\frac{1}{2}\mathbf{w}^T\mathbf{w}$$

$$(\text{조건}) \quad t^{(i)}(\mathbf{w}^T\mathbf{x}^{(i)} + b) \geq 1$$

- $t^{(i)}$는 i 번째 샘플의 클래스(양성/음성)

$$t^{(i)} = \begin{cases} -1 & x^{(i)}\text{가 음성인 경우} \\ 1 & x^{(i)}\text{가 양성인 경우} \end{cases}$$

조건식의 의미

$$(\text{조건}) \quad t^{(i)}(\mathbf{w}^T\mathbf{x}^{(i)} + b) \geq 1$$

- $\mathbf{x}^{(i)}$가 양성인 경우
  - $t^{(i)} = 1$
  - 따라서 $\mathbf{w}^T\mathbf{x}^{(i)} + b \geq 1$, 즉 양성으로 예측해야 함.
- $\mathbf{x}^{(i)}$가 음성인 경우
  - $t^{(i)} = -1$
  - 따라서 $\mathbf{w}^T\mathbf{x}^{(i)} + b \leq -1$, 즉 음성으로 예측해야 함.

소프트 마진 선형 SVM 분류기의 목표

$$\frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=0}^{m-1}\zeta^{(i)}$$

$$(\text{조건}) \quad t^{(i)}(\mathbf{w}^T\mathbf{x}^{(i)} + b) \geq 1 - \zeta^{(i)}$$

- $\zeta^{(i)} \geq 0$: **슬랙 변수**. i 번째 샘플에 대한 마진 오류 허용 정도 지정. (ζ는 그리스어 알파벳이며 '체타(zeta)'라고 발음함.)
- C: 아래 두 목표 사이의 트레이드오프를 조절하는 하이퍼파라미터
    - 목표 1: 결정 경계 도로의 폭을 가능하면 크게 하기 위해 $\|\mathbf{w}\|$ 값을 가능하면 작게 유지.
    - 목표 2: 마진 오류 제한. 슬랙 변수의 값 작게 유지.

ζ의 역할

- $\zeta^{(i)} > 0$이면 특정 샘플 $\mathbf{x}^{(i)}$에 대해 다음이 성립하여 마진 오류가 될 수 있음.

$$1 - \zeta^{(i)} \le t^{(i)}(\mathbf{w}^T\mathbf{x}^{(i)} + b) < 1$$

- 이유: 결정 경계면(하이퍼플레인) 상에서 보면 결정 함숫값이 1보다 작은 샘플이기에 실제 데이터셋의 공간에서는 결정 경계 도로 안에 위치하게 됨.

`C`와 마진 폭의 관계 (1부)

$$\frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=0}^{m-1}\zeta^{(i)}$$

$$(\text{조건})\quad t^{(i)}(\mathbf{w}^T\mathbf{x}^{(i)} + b) \geq 1 - \zeta^{(i)}$$

- 가정: 보다 간단한 설명을 위해 편향 b는 0이거나 무시될 정도로 작다고 가정. (표준화 전처리를 사용하면 됨.)
- C가 매우 큰 경우
  - ζ는 0에 매우 가까울 정도로 아주 작아짐.
  - 예를 들어 양성 샘플 $\mathbf{x}^{(i)}$에 대해, 즉 $t^{(i)} = 1$, $\mathbf{w}^T\mathbf{x}^{(i)}$ 가 1보다 크거나 아니면 1보다 아주 조금만 작아야 함. 즉, 결정 경계면의 기울기 $\|w\|$가 어느 정도 커야 함.
  - 결정 경계의 도로폭이 좁아짐.

`C`와 마진 폭의 관계 (2부)

$$\frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=0}^{m-1}\zeta^{(i)}$$

$$(\text{조건})\quad t^{(i)}(\mathbf{w}^T\mathbf{x}^{(i)} + b) \geq 1 - \zeta^{(i)}$$

- C가 매우 작은 경우
    - ζ가 어느 정도 커도 됨.
    - $\mathbf{w}^T\mathbf{x}^{(i)}$ 가 1보다 많이 작아도 됨. 즉, $\|w\|$ 가 작아도 됨.
    - 결정 경계의 도로폭이 넓어짐.