

5장 서포트 벡터 머신 2부

감사의 글

자료를 공개한 저자 오렐리앙 제롱과 강의자료를 지원한 한빛아카데미에게 진심어린 감사를 전합니다.

1부 주요 내용

- 선형 SVM 분류
- 비선형 SVM 분류

2부 주요 내용

- SVM 회귀
- SVM 이론

5.3 SVM 회귀

SVM 분류 vs. SVM 회귀

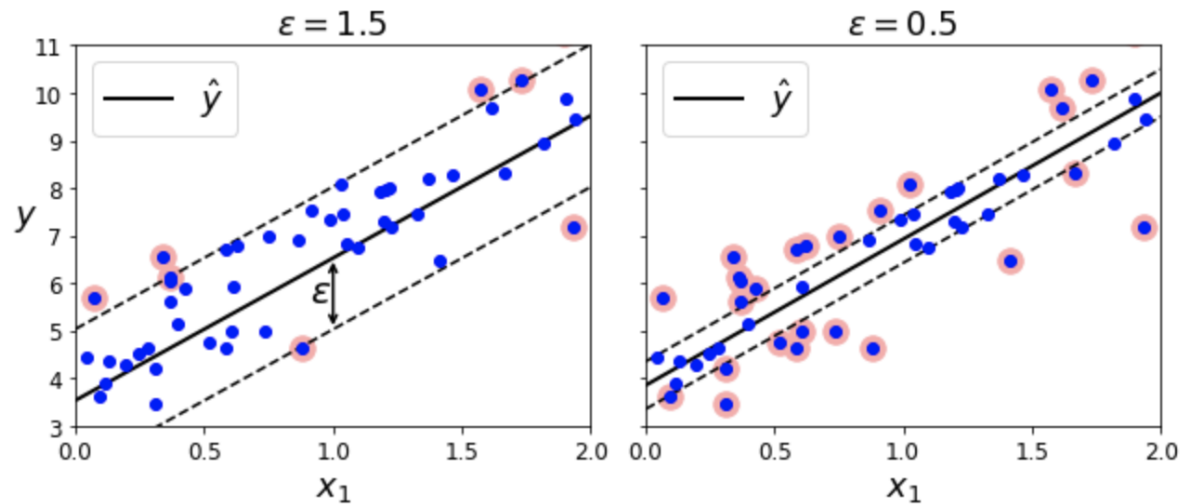
- SVM 분류
 - 목표: 마진 오류 발생 정도를 조절하면서 두 클래스 사이의 도로폭을 최대한 넓게 하기
 - 마진 오류: 도로 위에 위치한 샘플
- SVM 회귀
 - 목표: 마진 오류 발생 정도를 조절하면서 지정된 폭의 도로 안에 가능한 많은 샘플 포함하기
 - 마진 오류: 도로 밖에 위치한 샘플

선형 SVM 회귀

- 선형 회귀 모델을 SVM을 이용하여 구현
- 예제: LinearSVR 활용. `epsilon` 은 도로폭 결정

```
from sklearn.svm import LinearSVR  
svm_reg = LinearSVR(epsilon=1.5)
```

- 마진 안에 포함되는 샘플을 추가해도 예측에 영향 주지 않음. 즉 `epsilon` 에 둔감함.

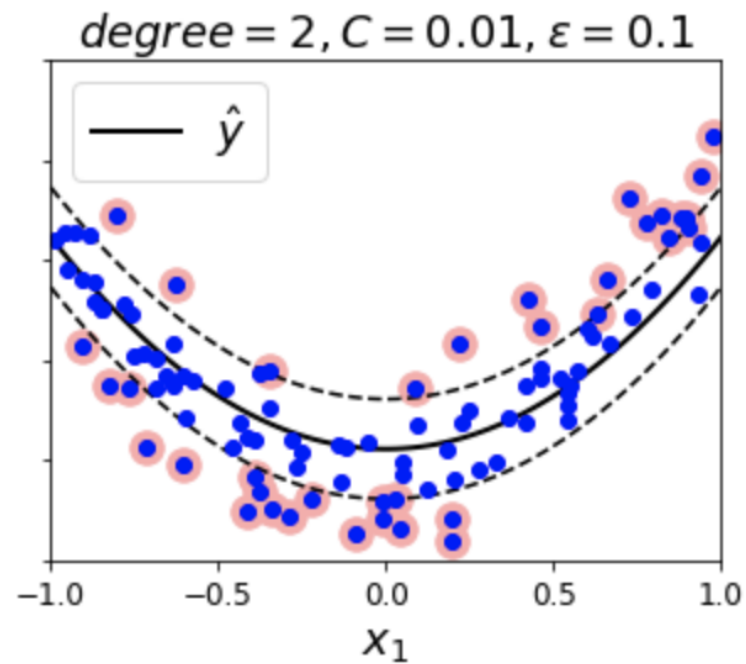
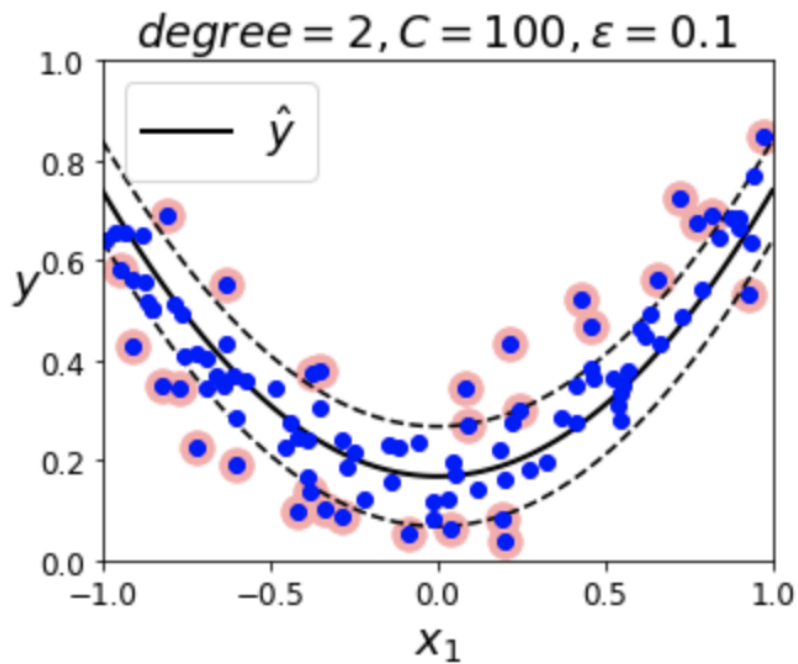


비선형 SVM 회귀

- SVC와 동일한 커널 트릭을 활용하여 비선형 회귀 모델 구현
- 예제: SVR + 다항 커널

```
# SVR + 다항 커널
from sklearn.svm import SVR

svm_poly_reg = SVR(kernel="poly", degree=2, C=100, epsilon=0.1, gamma="scale")
```

왼편 그래프	오른편 그래프
2차 다항 커널	2차 다항 커널
$C=100$ (규제 보다 적음)	$C=0.01$ (규제 보다 많음)
샘플에 더 민감	샘플에 덜 민감
도록뚱을 보다 좁게	도로뚱을 보다 넓게

회귀 모델 시간 복잡도

- LinearSVR: LinearSVC 의 회귀 버전
 - 시간 복잡도가 훈련 세트의 크기에 비례해서 선형적으로 증가
- SVR: SVC 의 회귀 버전
 - 훈련 세트가 커지면 매우 느려짐

5.4 SVM 이론

- (선형) SVM 작동 원리
 - 결정 함수와 예측
 - 목적 함수
 - 2차 계획법(QP, quadratic programming)
 - 쌍대 문제

- 커널 SVM 작동원리
 - 쌍대 문제를 해결할 때 커널 기법 활용 가능

- 온라인 SVM
 - 온라인 선형 SVM
 - 온라인 커널 SVM

(선형) SVM 작동 원리: 결정 함수와 예측

선형 SVM 분류기 모델의 결정 함수

$$\begin{aligned}h(\mathbf{x}) &= \mathbf{w}^T \mathbf{x} + b \\&= w_1 x_1 + \cdots + w_n x_n + b\end{aligned}$$

선형 SVM 분류기 예측

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

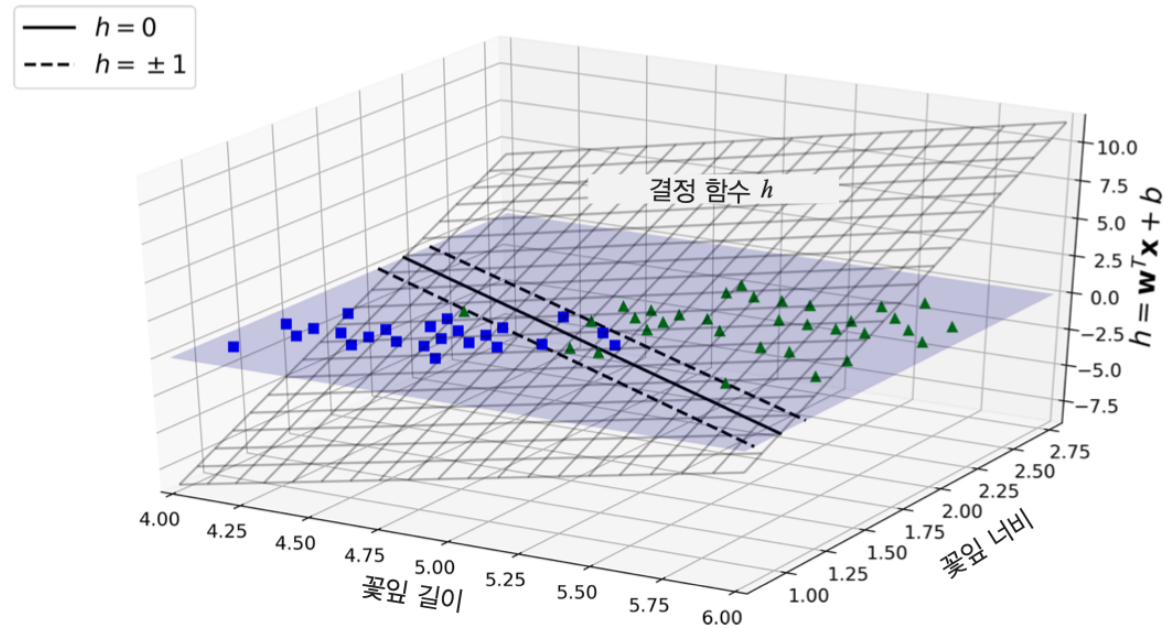
결정 경계

- 결정 함수의 값이 0인 점들의 집합

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

- 결정 경계 예제

- 붓꽃 분류: 꽃잎 길이와 너비를 기준으로 Iris-Virginica(초록색 삼각형) 품종 여부 판단



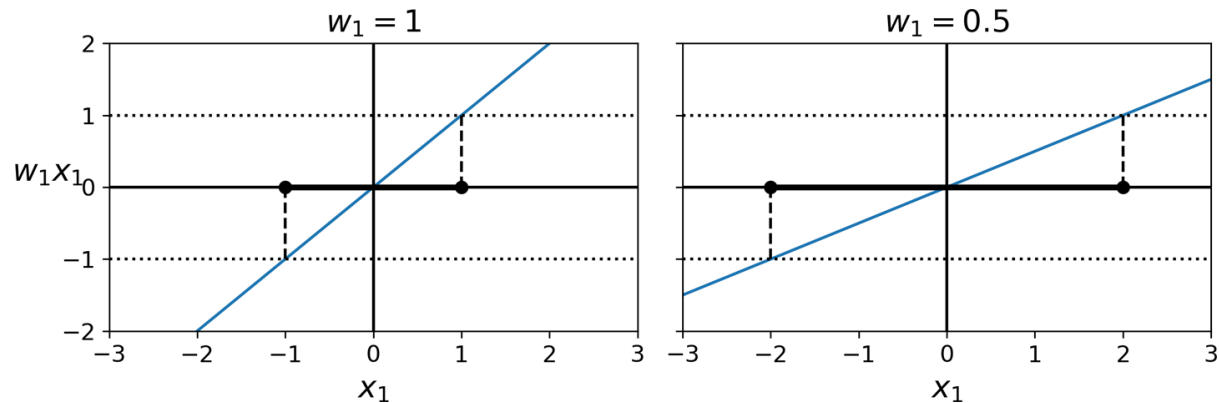
- 두 점선에 유의할 것

- $h(\mathbf{x})$ 가 1 또는 -1인 샘플들의 집합
- 마진과 밀접하게 관련됨.

(선형) SVM 작동 원리: 목적 함수

결정 함수의 기울기와 마진 폭

- 결정 함수의 기울기가 작아질 수록 마진 폭이 커짐. 아래 그림 참조
- 결정 함수의 기울기가 $\|\mathbf{w}\|$ 에 비례함.



- 마진을 크게 하기 위해 $\|\mathbf{w}\|$ 를 최소화 해야 함.
 - 하드 마진: 모든 양성(음성) 샘플에 대한 결정 함수의 값이 1(-1)보다 크다(작다)
 - 소프트 마진: 모든 샘플에 대한 결정 함수의 값이 지정된 값 이상 또는 이하이어야 한다.

하드 마진 선형 SVM 분류기의 목적 함수

- 목적 함수:

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

- 아래 조건 하에서 목적 함수를 최소화 시키는 \mathbf{w} 와 b 를 구해야 함:

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

- 단, 다음이 성립:

- $x^{(i)}$: i 번째 샘플
- $t^{(i)}$: 양성 샘플일 때 1, 음성 샘플일 때 -1

소프트 마진 선형 SVM 분류기의 목적 함수

- 목적 함수:

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

- 아래 조건 하에서 목적 함수를 최소화 시키는 \mathbf{w} 와 b 를 구해야 함:

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

- 단, 다음이 성립:
 - $x^{(i)}$: i 번째 샘플
 - $t^{(i)}$: 양성 샘플일 때 1, 음성 샘플일 때 -1
 - $\zeta^{(i)} \geq 0$: **슬랙 변수**. i 번째 샘플에 대한 마진 오류 허용 정도 지정.
- C : 아래 두 목표 사이의 트레이드오프를 조절하는 하이퍼파라미터
 - 목표 1: 슬랙 변수의 값을 작게 만들기
 - 목표 2: 마진을 크게 하기 위해 $\frac{1}{2} \mathbf{w}^T \mathbf{w}$ 값을 가능하면 작게 만들기

(선형) SVM 작동 원리: 2차 계획법(QP)

- 하드(소프트) 마진 문제: 선형 제약조건이 있는 블록 2차 최적화 문제
- 2차 계획법(QP, quadratic programming) 문제로 알려짐.
- 해법에 대한 설명은 이 책의 수준을 벗어남.

(선형) SVM 작동 원리: 쌍대 문제

- 쌍대 문제(dual problem): 주어진 문제의 답과 동일한 답을 갖는 문제
- 하드(소프트) 마진과 관련된 2차 계획법 문제의 답을 보다 쉽게 해결할 수 있는 쌍대 문제를 이용하여 해결 가능

선형 SVM 목적 함수의 쌍대 문제

- 아래 식을 최소화하는 α 찾기. 단, $\alpha^{(i)} > 0$:

$$\frac{1}{2} \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} \alpha^{(i)} \alpha^{(j)} t^{(i)} t^{(j)} \mathbf{x}^{(i)T} \mathbf{x}^{(j)}$$

$$\sum_{i=0}^{m-1} \alpha^{(i)}$$

- 쌍대 문제의 답 $\hat{\alpha}$ 를 이용하여 $\hat{\mathbf{w}}$ 와 \hat{b} 를 선형 SVM 모델의 파라미터로 활용
 - n_s : 서포트 벡터 수, 즉, $\hat{\alpha}^{(i)} > 0$ 인 샘플 수

$$\hat{\mathbf{w}} = \sum_{i=0}^{m-1} \hat{\alpha}^{(i)} t^{(i)} \mathbf{x}^{(i)}$$

$$\hat{b} = \frac{1}{n_s} \sum_{i=0, \hat{\alpha}^{(i)} > 0}^{m-1} (t^{(i)} - \hat{\mathbf{w}}^T \mathbf{x}^{(i)})$$

커널 SVM 작동 원리

쌍대 문제와 커널 SVM

- 커널 SVM이 작동 원리는 원래의 문제가 아닌 쌍대 문제 해결과 관련됨.

- 특히 아래 쌍대 목적 함수에서 사용된 $\mathbf{x}^{(i)T} \mathbf{x}^{(j)}$ 에 주의해야 함.

$$\frac{1}{2} \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} \alpha^{(i)} \alpha^{(j)} t^{(i)} t^{(j)} \boxed{\mathbf{x}^{(i)T} \mathbf{x}^{(j)}} - \sum_{i=0}^{m-1} \alpha^{(i)}$$

예제: 2차 다항 커널 작동 아이디어

- 원래 아래 2차 다항식 함수를 적용한 후에 쌍대 목적 함수의 최적화 문제를 해결해야 함.

$$\phi(\mathbf{x}) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)^T$$

- 원래 아래 식의 최적화 문제를 해결해야 함.

$$\frac{1}{2} \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} \alpha^{(i)} \alpha^{(j)} t^{(i)} t^{(j)} \boxed{\phi(\mathbf{x}^{(i)})^T \phi(\mathbf{x}^{(j)})} - \sum_{i=0}^{m-1} \alpha^{(i)}$$

- 하지만 다음이 성립함

$$\phi(\mathbf{a})^T \phi(\mathbf{b}) = (\mathbf{a}^T \mathbf{b})^2$$

- 따라서 2차 다항식 함수 ϕ 전혀 적용할 필요 없이 아래 함수에 대한 최적화 문제를 해결하면 됨.

$$\frac{1}{2} \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} \alpha^{(i)} \alpha^{(j)} t^{(i)} t^{(j)} \boxed{(\mathbf{x}^{(i)T} \mathbf{x}^{(j)})^2} - \sum_{i=0}^{m-1} \alpha^{(i)}$$

- 커널 기법으로 구해진 쌍대문제의 해 $\hat{\alpha}$ 를 이용하여 예측값 $h(\phi(\mathbf{x}))$ 또한 $\phi(\mathbf{x})$ 없이 계산할 수 있음.

예제: 지원되는 커널

- 선형:

$$K(\mathbf{a}, \mathbf{b}) = \mathbf{a}^T \mathbf{b}$$

- 다항식:

$$K(\mathbf{a}, \mathbf{b}) = (\gamma \mathbf{a}^T \mathbf{b} + r)^d$$

- 가우시안 RBF:

$$K(\mathbf{a}, \mathbf{b}) = \exp(-\gamma \|\mathbf{a} - \mathbf{b}\|^2)$$

- 시그모이드:

$$K(\mathbf{a}, \mathbf{b}) = \tanh(\gamma \mathbf{a}^T \mathbf{b} + r)$$

온라인 SVM

- 온라인 학습: 새로운 샘플에 대해 점진적으로 학습하기

선형 온라인 SVM

- 특정 비용함수를 최소화하기 위한 경사하강법 사용
- 예제: 사이킷런의 SGDClassifier
 - `loss` 하이퍼파라미터를 `hinge` 로 설정하면 선형 SVM 모델 지정

비선형 온라인 SVM

- 온라인 커널 SVM 구현 가능.
- 하지만 신경망 알고리즘 사용 추천