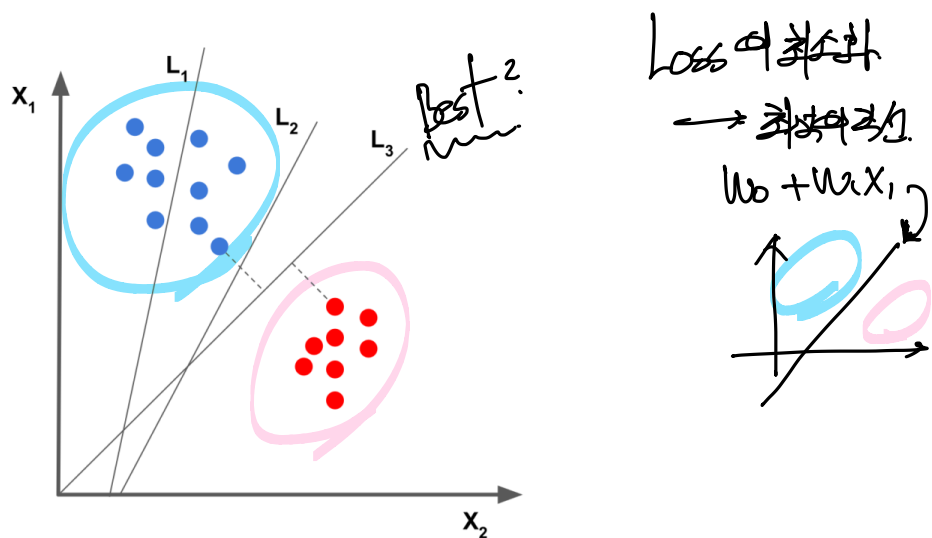


# 서포트 벡터 머신

## 도입

- 아래와 같은 데이터 셋이 주어졌을때 어떻게 분류 해야 할까?
  - 하나의 선형모델을 찾고 서로다른 Class 의 Sample을 분리한다.
  - 하지만 여러개의 선형모델이 존재한다면 어떻게 더 좋은 방법일까?
  - $L_1, L_2, L_3$  중 어느 직선을 사용해야 할까?



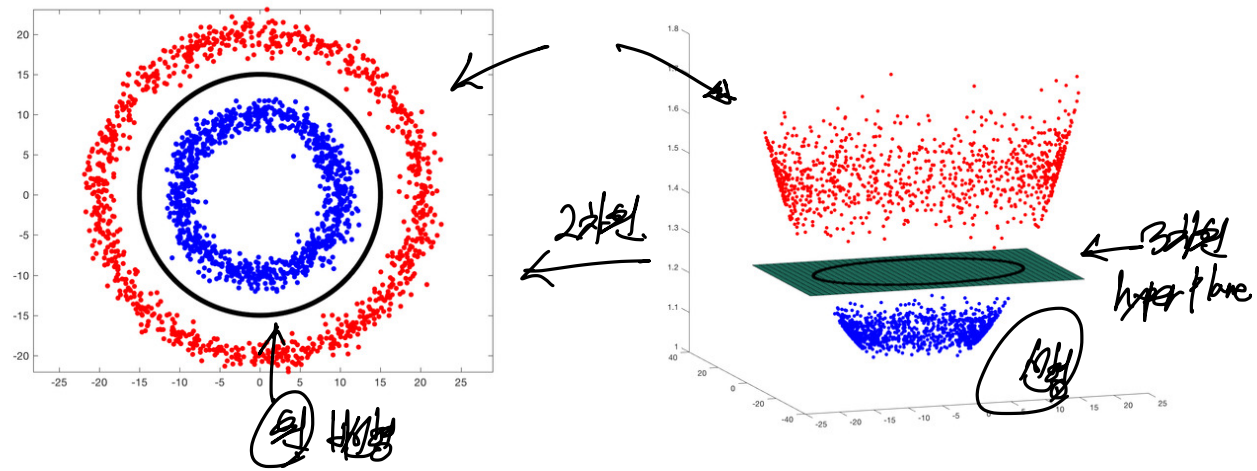
## Question

- Model** → **Training(&Generalization)** 시 데이터에 예측 대한 성능이 좋아져야 한다.
- 하지만 TrainingData 에 대한 성능이 높아지면 Generalization 성능은 낮아지게 된다. 이를 해결 할수 없을까?

??  
test set 성능

## 분류문제

- 기본적인 분류문제 : 3차원, 분류, 선형



- 2차원에서 두개의 클래스를 가장 잘 분류한 직선은 무엇일까?

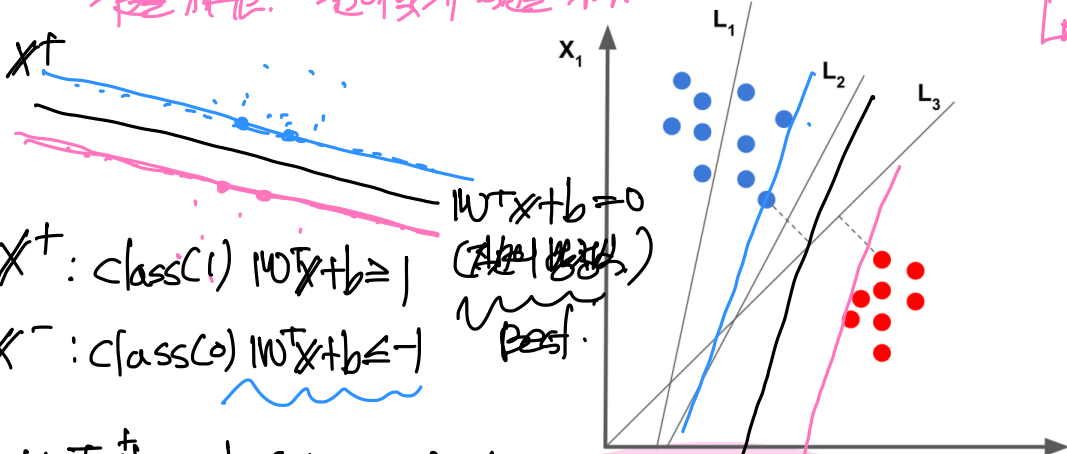
- 두개의 Class를 나누는 평면은 무수히 많지만 어떻게 나눠야 최적화된 평면이 만들어질까? (아름답)

- 좋은 기준은 무엇인가?

가려운 사람. 가려워맞게 잘잘과.

$$\text{Loss}(G(W^T X), y) \text{ 최소화}$$

↓ 알라딘과 함께



$$X^+ : \text{class}(1) \quad W^T X + b \geq 1$$

$$X^- : \text{class}(0) \quad W^T X + b \leq -1$$

$$|W^T X + b| = 1 \text{ (최적화 조건) 이고 } X^+ = X^- + \lambda W \text{ (변환선에서 이동)}$$

$$W^T (X^- + \lambda W) + b = 1 \Rightarrow W^T X^- + \lambda W^T W + b = 1$$

Sample Space

$$\lambda W^T W = |+| = 2$$

- Sample Space 에서의 분할 HyperPlane 은  $W^T X + b = 0$  으로 표현 할수 있다.

$$\|W\|_2 = \sqrt{W^T W} = \frac{2}{\lambda} \quad \text{Note: } \|W\|_p = (\sum w_i^p)^{1/p} = \sqrt{w_1^2 + w_2^2 + \dots}$$

## 1 선형 SVM 분류 $\lambda = \frac{2}{W^T W}$

- 라지마진 분류 (maximizing margin over the training set)

- SVM 에서 좋은 기준은 Margin 이다.

- Margin 을 최대화 시키는 Hyperplane을 찾는다.

- Minimizing generalization error (=testing error)  $\Rightarrow$  좋은 예측 성능을 나타낸다.

$$W = (w_1, w_2, \dots, w_n)$$

$$W^T W = [w_1, w_2, \dots, w_n] \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = w_1^2 + w_2^2 + \dots + w_n^2$$

- Margin 이 뭐야?

- 각 class에서 가장 가까운 관측치 사이의 거리를 나타 낸다.

- Margin 은  $W$ (기울기) 로 표현 가능

$$\text{Margin} = \text{distance}(X^+, X^-)$$

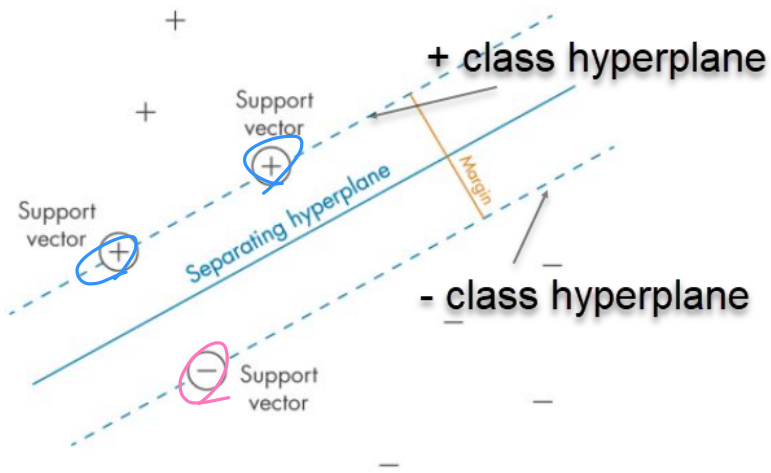
$$= \|X^+ - X^-\|_2$$

$$= \|(X^- + \lambda W) - X^+\|_2 = \|\lambda W\|_2$$

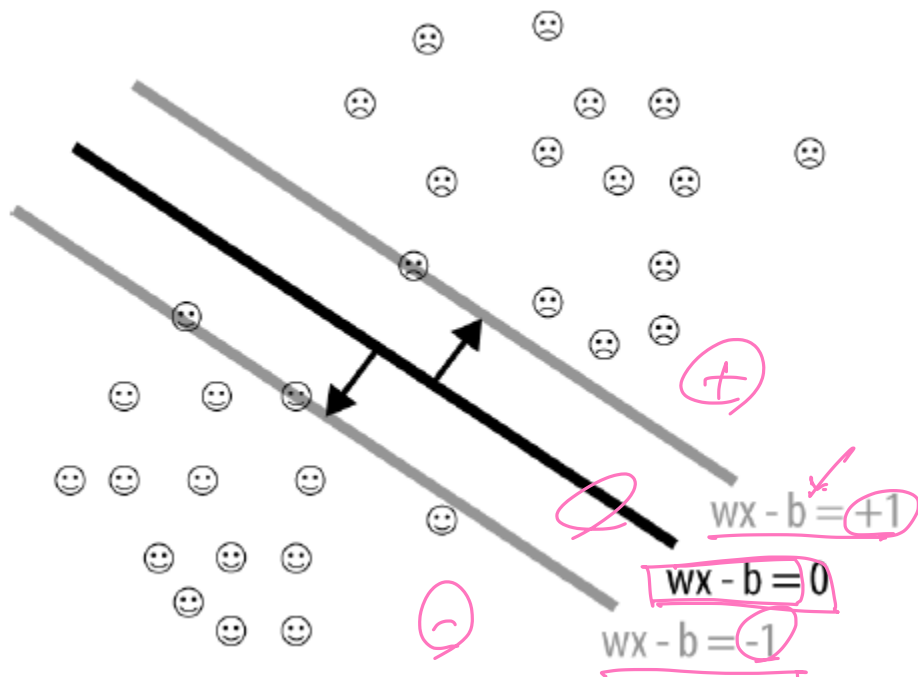
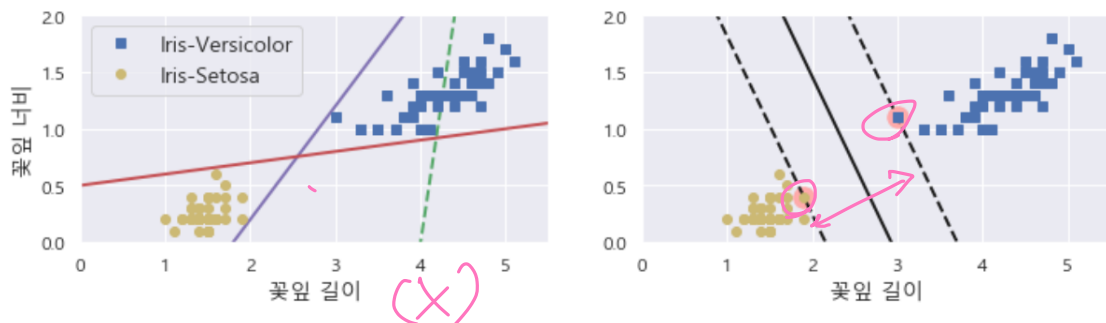
$$\lambda = \frac{2}{W^T W} \Rightarrow \sqrt{W^T W} = \frac{2}{\|W\|_2} = \frac{2}{\|W\|_2}$$

$$\text{결론: Margin} = \frac{2}{\|W\|_2}$$

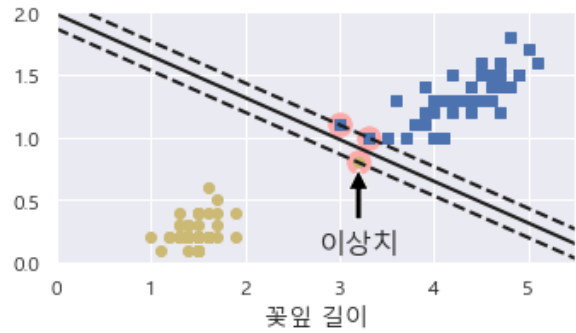
$$\max \text{margin} = \max \frac{2}{\|W\|_2} = \min \frac{\|W\|_2}{2} \text{ 이 } W \text{ 는 } W^T X_i + b \geq 1$$



- 경계에 위치한 샘플을 **서포트 벡터** 라고 한다.
- 오른쪽 그래프에서 붉은색으로 표시된 점, 즉 데이터 포인트를 **Support Vector** 라고 하는데 그 이유는 Margin이 이 붉은색으로 표시된 데이터에 의해 결정되기 때문이다.



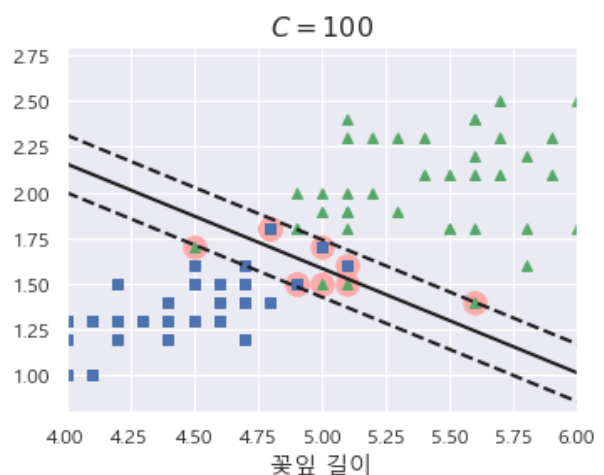
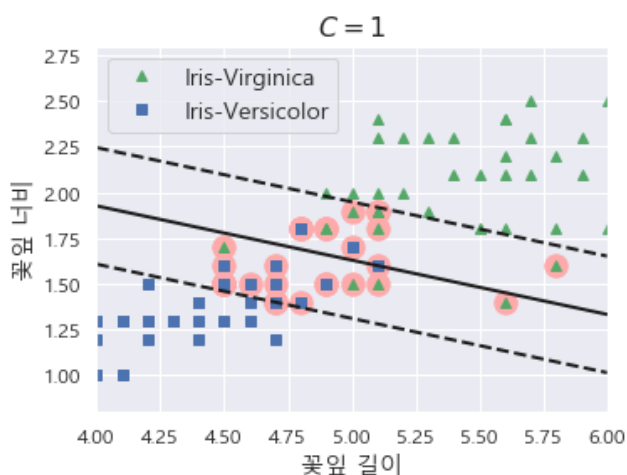
## 하드마진 vs 소프트 마진



- 하드마진 방식은 매우 엄격하게 두 개의 클래스를 분리하는 분리초평면(위에서는 선형분리)을 구하는 방법이다
- 모든 Train Set은 분리 초평면을 사이에 두고 무조건 한 클래스에 속해야 한다.
- 하지만, 몇 개의 **노이즈** 또는 이상치로 인해 두 클래스를 분류하는 분리 초평면을 구할 수조차 없거나(위의 왼쪽 그래프), 제대로 구하지 못할 수도 있다(위의 오른쪽 그래프).
- 하드 마진 분류 : 모든 샘플이 쪽으로 올바르게 분류되어 있는 경우
  - 왼쪽 그래프 : 하드 마진을 찾을수 없는 경우
  - 오른쪽 그래프 : 결정 경계에 이상치가 없어 결정경계가 매우 일반화 되어 있는 경우

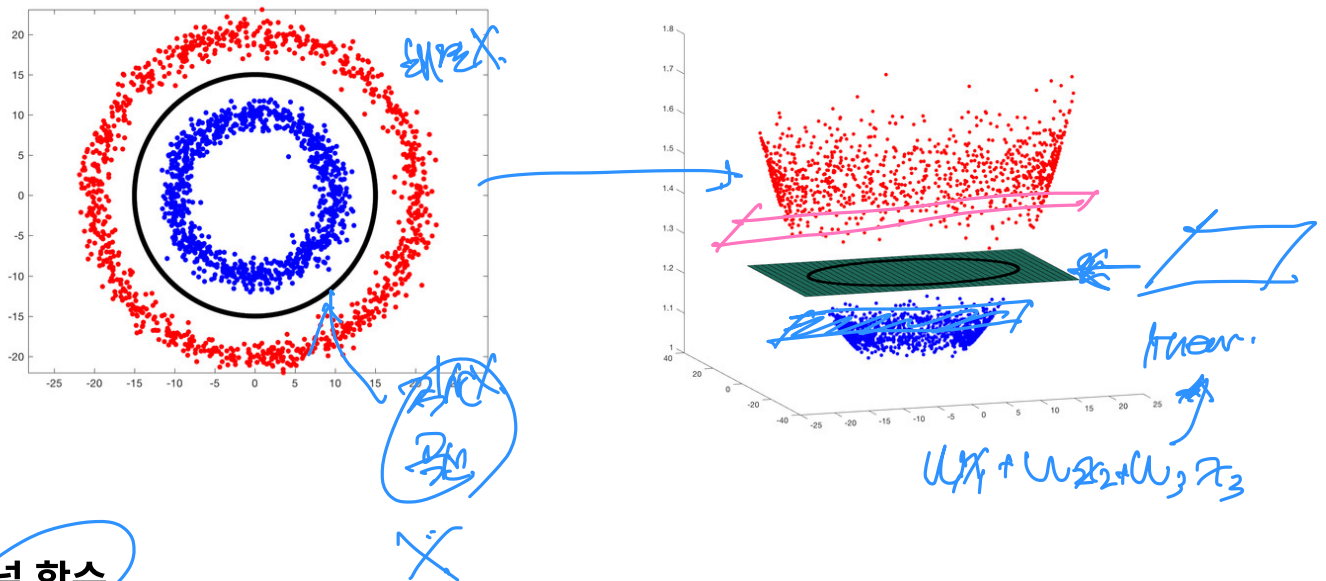
## 소프트 마진

- **소프트 마진** SVM은 기본적으로 하드 마진 방법을 기반으로 하는데, 가장 큰 차이점은 Support Vectors가 위치한 경계선에 약간의 **여유 변수**(Slack Variable)을 두는 것
  - Scikit-Learn의 SVM 모델에서는 이러한 여유 변수를 **C** 라는 하이퍼파라미터를 제공\
  - **C** 는 일종의 **penalty**라고 볼 수 있다.
    - **C** 값을 줄이면 오류를 허용하는 정도가 커지며, Margin 또한 커진다.
    - **C** 값을 크게하면 오류를 허용하는 정도가 작아지며, Margin 또한 작아진다.



## 2 비선형 SVM 분류

- 비선형 데이터 셋을 다루는 방법?
  - 더 높은 차원의 특성 공간으로 투영하여 특성공간 내에서 선형 분리 가능하게 만들수 있다
$$x \mapsto \phi(x)$$



## 커널 함수

- $\phi(x)$ 를  $x$ 를 투영시킨후의 고유 벡터 라고 한다면, 특성 공간에서 분할 초평면에 대응하는 모델은 다음과 같이 표현 가능하다

$$f(x) = W^T \phi(x) + b$$

2차원  $\rightarrow$  2차원.

- 다항특성과 같은 특성을 더 추가 하는방법으로 해결 가능하다.
  - 다항식 커널
  - 유사도 특성
  - 가우시안 RBF 커널
  - 계산 복잡도

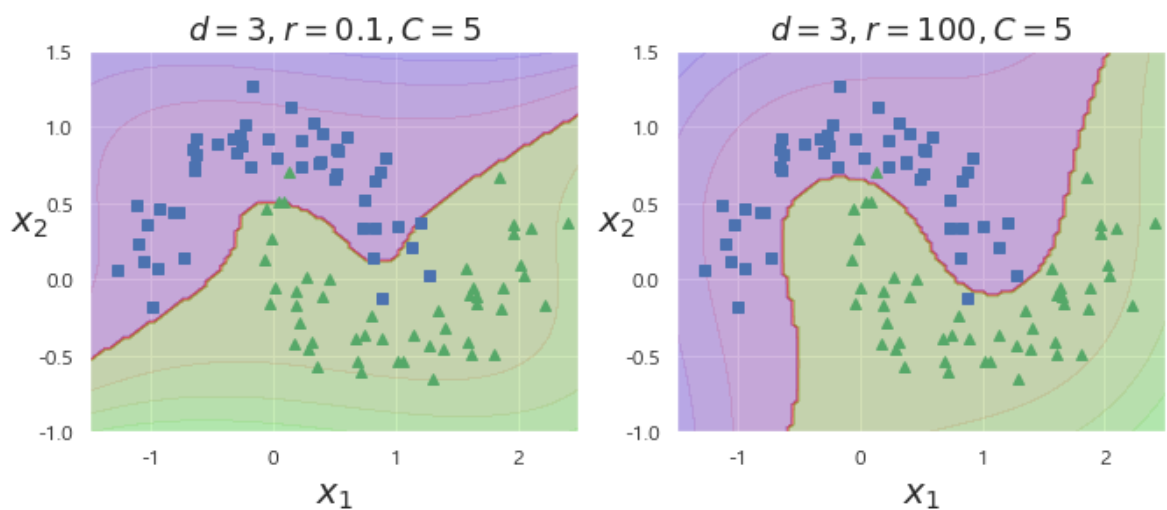
파이썬 클래스	시간 복잡도	외부 메모리 학습 지원	스케일 조정의 필요성	커널 트릭
LinearSVC	$O(m \times n)$	아니오	예	아니오
SGDClassifier	$O(m \times n)$	예	예	아니오
SVC	$O(m^2 \times n) \sim O(m^3 \times n)$	아니오	예	예

▲ 표 5-1 SVM 분류를 위한 사이킷런 파이썬 클래스 비교

## 다항식 커널

- 다항식 특성을 추가하는 것은 간단하고 (SVM뿐만 아니라) 모든 머신러닝 알고리즘에서 잘 작동한다.

- 하지만 낮은 차수의 다항식은 매우 복잡한 데이터셋을 잘 표현하지 못하고 높은 차수의 다항식은 굉장히 많은 특성을 추가하므로 모델을 느리게 만든다.



## 유사도 특성

비선형 특성을 다루는 또 다른 기법은 각 샘플이 특정 랜드마크(landmark)와 얼마나 닮았는지 측정하는 유사도 함수(similarity function)로 계산한 특성을 추가하는 것

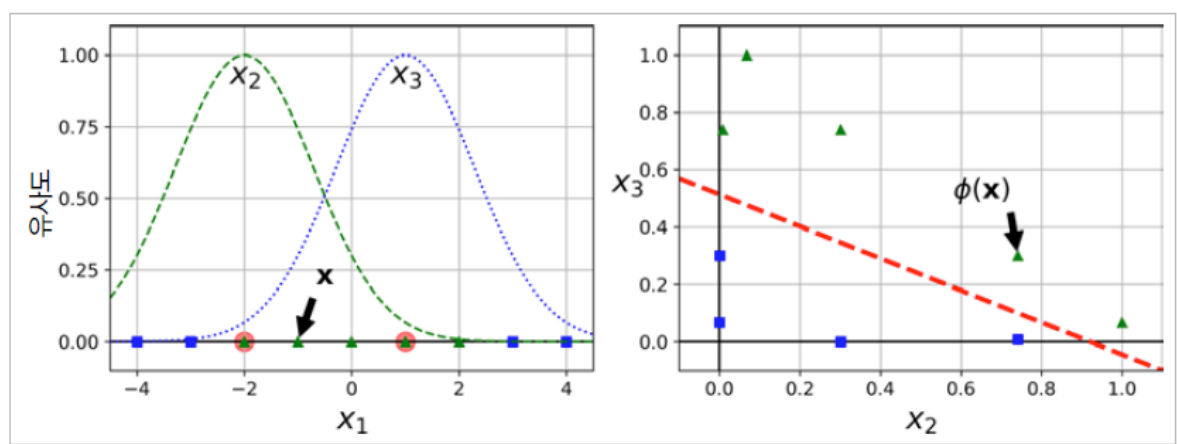
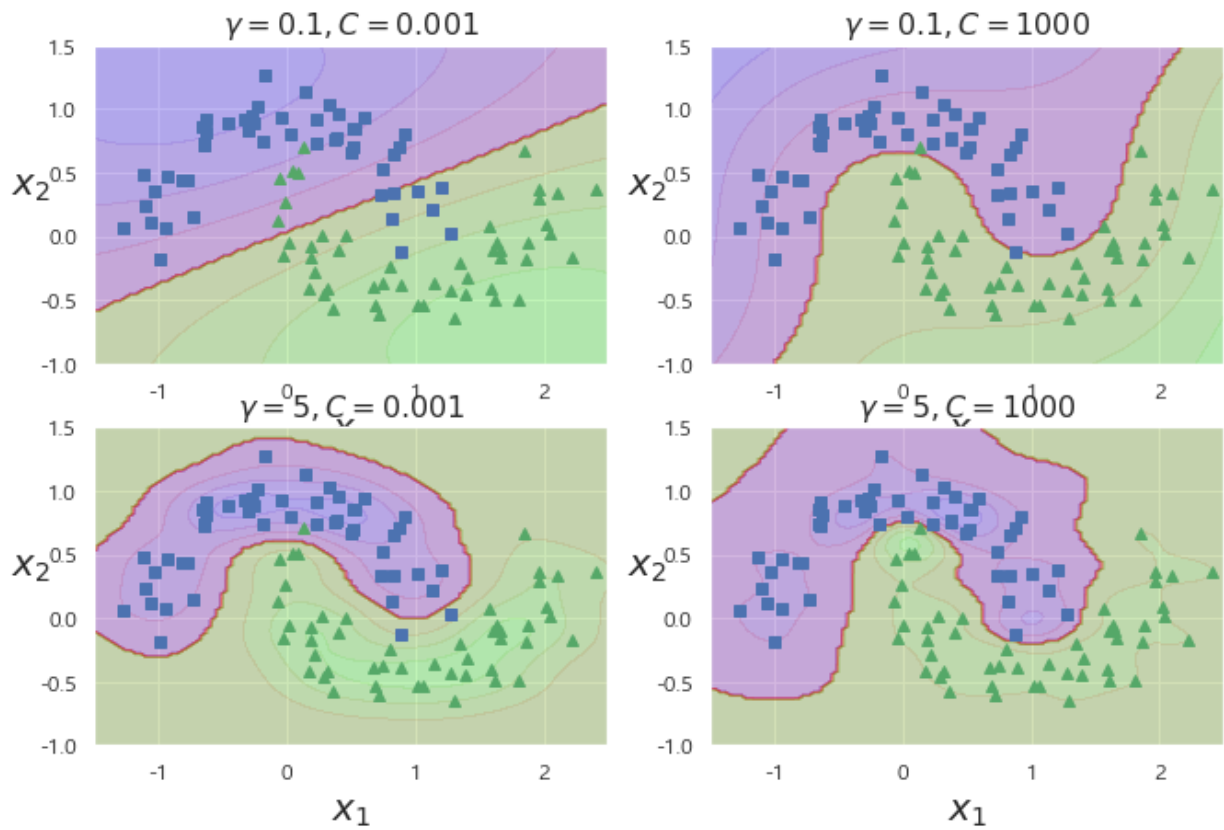


그림 5-8 가우시안 RBF를 사용한 유사도 특성

*learn 가변 변환!*

## 가우시안 RBF 커널

- PolynomialFeatures 변환기를 이용해 차원 지정하여 코드를 작성한다.



## 계산 복잡도

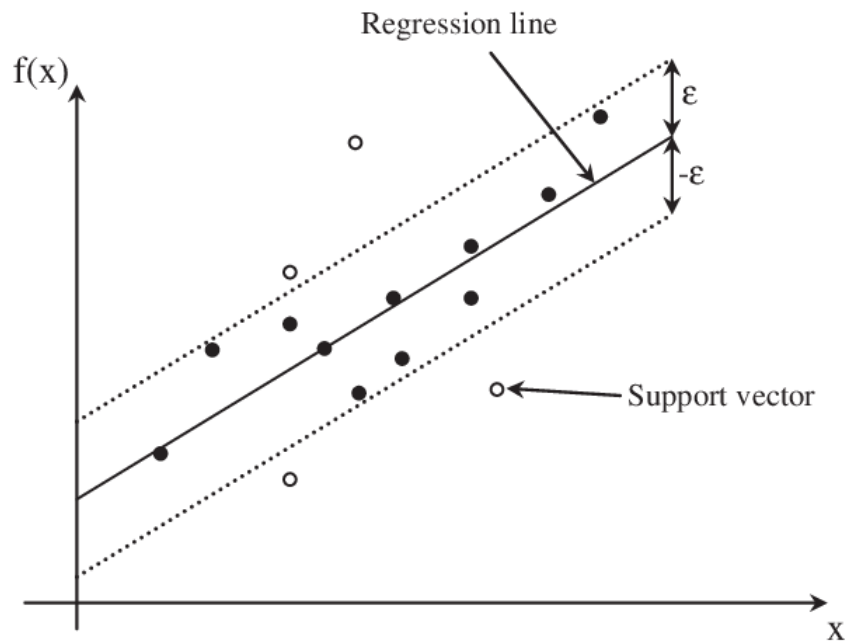
파이썬 클래스	시간 복잡도	외부 메모리 학습 지원	스케일 조정의 필요성	커널 트릭
LinearSVC	$O(m \times n)$	아니오	예	아니오
SGDClassifier	$O(m \times n)$	예	예	아니오
SVC	$O(m^2 \times n) \sim O(m^3 \times n)$	아니오	예	예

▲ 표 5-1 SVM 분류를 위한 사이킷런 파이썬 클래스 비교

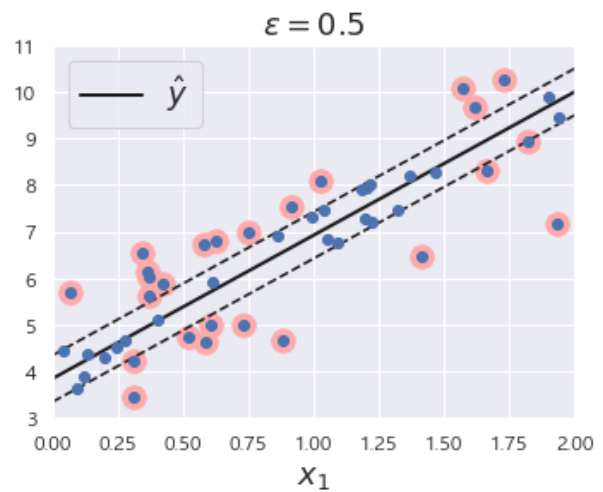
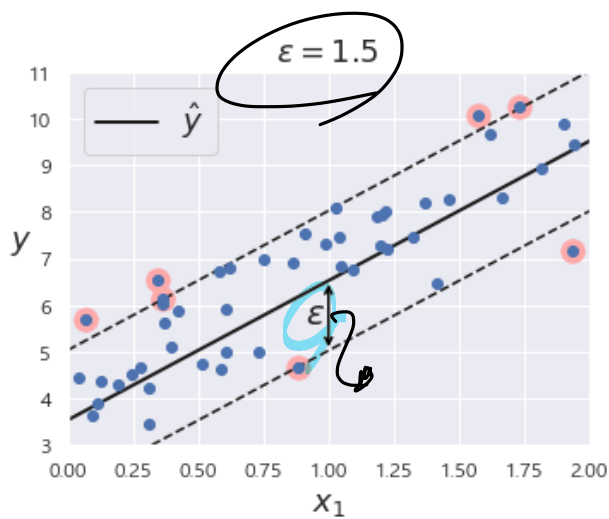
## 3 Support Vector Regression



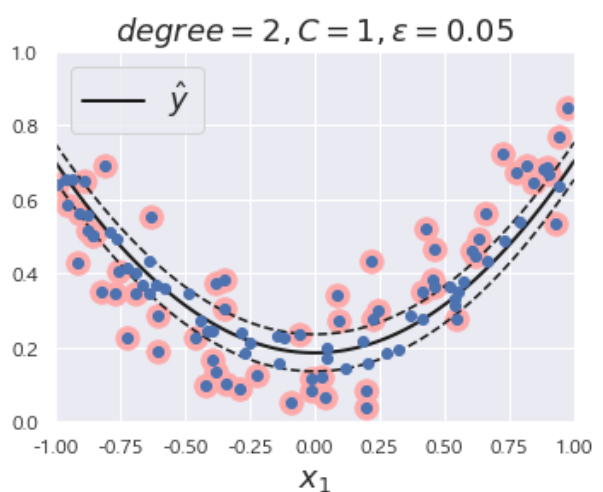
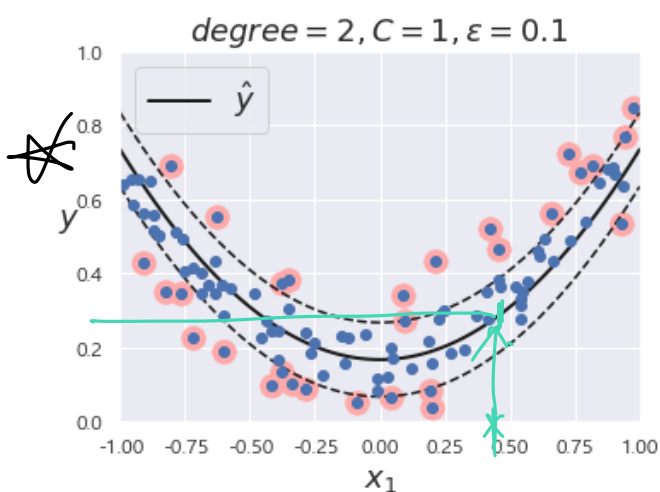




- SVM 은 선형, 비선형 분류 뿐만 아니라 선형, 비선형 회귀에도 사용할 수 있다.
- SVM을 회귀에 적용하는 방법은 기존 SVM 의 방식을 반대로 적용하면 된다.
  - 마진 오류 안에서 두 클래스 간의 폭이 최대가 되도록한다.
  - SVM 회귀는 제한된 마진오류(도로 밖의 샘플) 안에서 도로 안에 가능한 한 많은 샘플이 들어가도록 학습한다.
- 마진을 결정하고 최대한 Dataset이 많이 들어가게 만들어 준다.



## SVR을 이용한 비선형 SVM 회귀





## 추가학습

OCSVM

이성치 분류.