

5.2 비선형 SVM 분류

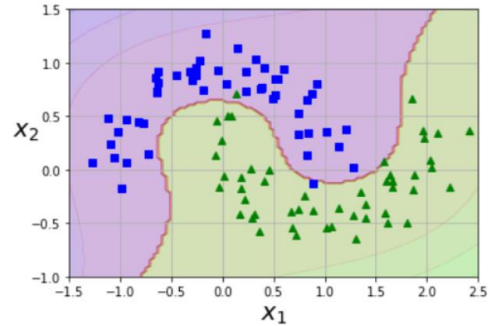
- 특성을 추가해 선형으로 구분되지 않는 것을 구분될 수 있게 하는 것
- 사이킷런을 사용해 구현하는 경우, PolynomialFeatures, StandardScaler, LinearSVC를 연결해 파이프라인을 만들 수 있다.

```
from sklearn.datasets import make_moons
from sklearn.preprocessing import PolynomialFeatures

X,y = make_moons(n_samples=100, noise=0.15, random_state=42)

polynomial_svm_clf = Pipeline([
    ('poly_features', PolynomialFeatures(degree=3)),
    ('scaler', StandardScaler()),
    ('svm_clf', LinearSVC(C=10, loss='hinge', max_iter=2000))
])

polynomial_svm_clf.fit(X,y)
```



1) 다항식 커널

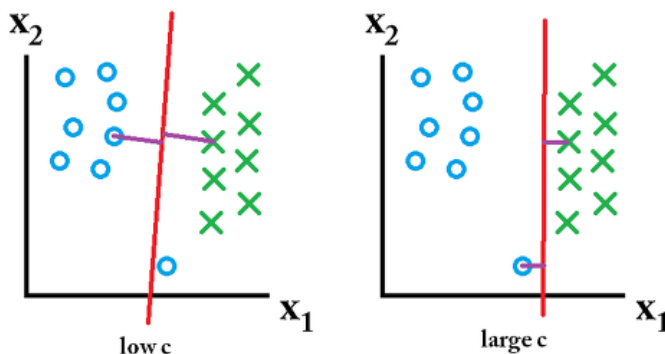
- 간단하고, 모든 ML알고리즘에서 잘 작동
- 낮은 차수 다항식은 복잡한 데이터셋 표현에 부족
- 높은 차수 다항식은 특성 추가로 인한 성능 저하 야기

* 커널 트릭 (실제로는 특성을 추가하지 않으면서 다항식 특성을 많이 추가한 듯한 결과)

```
from sklearn.svm import SVC
poly_kernel_svm_clf = Pipeline([
    ('scaler', StandardScaler()),
    ('svm_clf', SVC(kernel='poly', degree=3, coef0=1, C=5)),
    # coef0 parameter = 높은 차수와 낮은 차수에 얼마나 영향을 받을지 결정
])

poly_kernel_svm_clf.fit(X,y)
```

- Coef 매개변수는 모델이 높은(혹은 낮은) 차수에 받는 영향 조절
- Degree 매개변수는 커널 차수 표현 (지금은 3차)
- C(cost) 는 얼마나 많은 데이터 샘플이 다른 클래스에 놓이는 것을 허용하는지 결정 (작을수록 많이 허용 = 이상치가 있을 가능성을 크게 잡아 rough하게 결정경계 찾기)



※C의 값이
- 너무 낮음 -> 과소적합
- 너무 높음 -> 과대적합

※C의 값이
- 있음 -> 하드마진
- 없음 -> 소프트마진

5.2 비선형 SVM 분류

2) 유사도 함수로 계산한 특성 추가

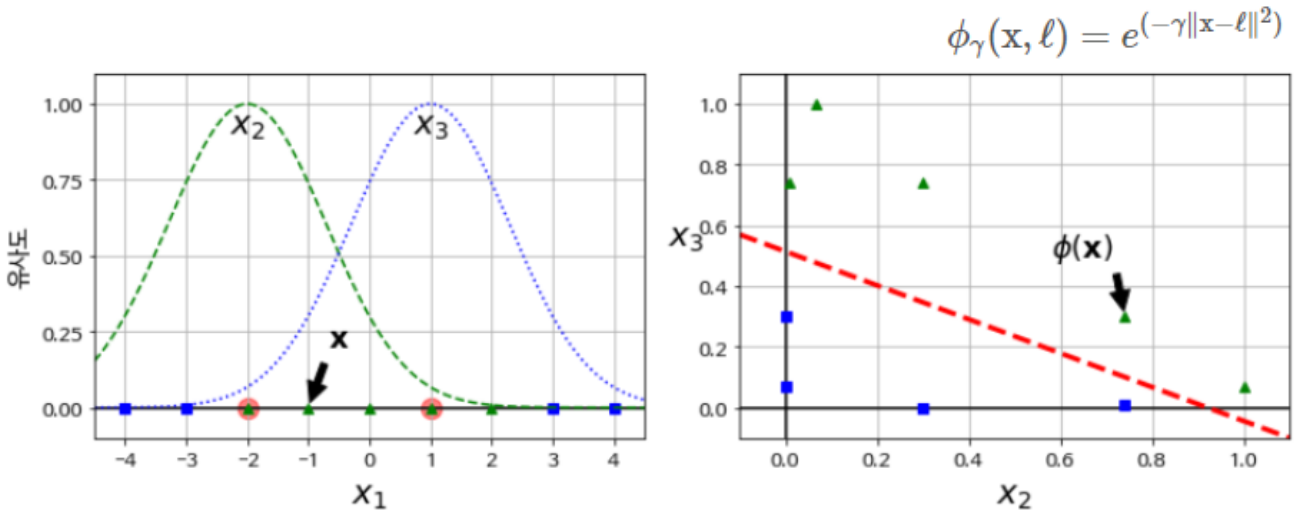
- 랜드마크와 샘플이 얼마나 닮았는지 측정하는 유사도 함수
- Ex) 앞 예시에 $x_1 = -2, x_2 = 1$ 를 추가하고 (랜드마크)
 $\gamma = 0.3$ 인 방사기저 함수를 유사도 함수로 정의해보자

* 방사기저함수 (좀 더 자세히 알아보자)

$$\phi_\gamma(\mathbf{x}, \ell) = e^{(-\gamma \|\mathbf{x} - \ell\|^2)}$$

$$K_{\text{RBF}}(\mathbf{x}, \mathbf{x}') = \exp[-\gamma \|\mathbf{x} - \mathbf{x}'\|^2]$$

- γ is a parameter that sets the "spread" of the kernel.
- $\gamma > 0$ 일 때, $-\gamma \|\mathbf{x} - \mathbf{x}'\|^2$ 는 커진다
- 가까운 벡터들이 더 큰 RBF 커널 값을 가지게 되면서 종모양 커브를 생성하게 된다.



- $x_1 = 1$ 샘플 보봐라, 첫 번째 랜드마크에서 1만큼 떨어져있고, 2번째에선 2만큼
- 새로 만든 특성은 $\exp(-0.3 * 1^2) \approx 0.74$ 량, $\exp(-0.3 * 2^2) \approx 0.30$
- 이렇게 방사함수에 있는 요소들로 선형 분류를 이끌어낼 수 있다.
- 랜드마크는 데이터셋의 모든 샘플 위치에 생성하면 차원이 커지며, 선형 분류될 가능성이 높다.
- 단, 훈련세트가 크다면 동일한 크기의 많은 특성이 생기게 된다.

5.2 비선형 SVM 분류

3) 가우시안 RBF 커널

커널 기법은 주어진 데이터를 고차원 특징 공간으로 사상해주는 것, 고차원 공간에 사상되면 원래의 차원에서는 보이지 않던 선형으로 분류가 가능하며, RBF커널 기법은 무한 차원으로 사상 (projection)될 수 있다.

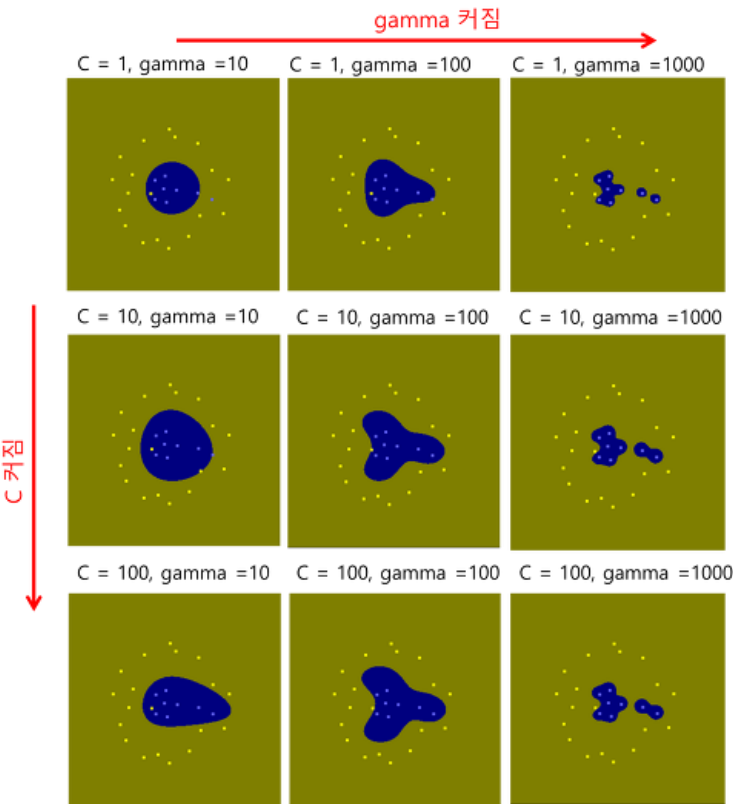
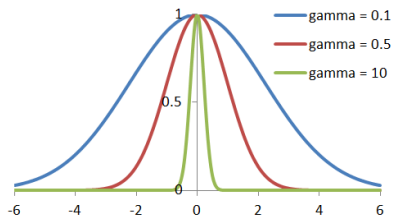
커널에는 Polynomial, Sigmoid 등이 있으나 가우시안RBF가 가장 성능이 좋아 자주 사용한다. 각 커널마다 매개변수가 있는데 RBF의 경우에는 gamma이며, SVM 기본 매개변수인 c도 있어서 2개의 매개변수를 설정해줘야 한다.

* 가우시안 RBF vs. 가우시안 RBF 커널? (좀 더 자세히 알아보자)

	가우시안 RBF	가우시안 RBF 커널
정규화	가중치의 제곱 규범에 따라 페널티 사용	매개변수화가 아니라 모델의 기능에 대한 페널티
	사용된 데이터 샘플에 따라 페널티 결정	샘플에 관계 없이 유도된 피쳐 공간 동일

[Gamma = 하나의 데이터 샘플이 영향력을 행사하는 거리 결정]

- Gamma를 증가시키면
 - 종모양 그래프가 좁아져서 각 샘플 영향범위가 작아진다.
 - 결정경계가 불규칙해지고, 샘플 따라 휘어진다.
- Gamma를 감소시키면
 - 넓은 종모양 그래프를 만들면서 결정경계가 부드러워짐



※gamma

- 결정경계 곡률 조정
- 데이터 포인트 영향 거리 down
- 너무 낮음-> 과소적합
- 너무 높음-> 과대적합

※ C의 값이

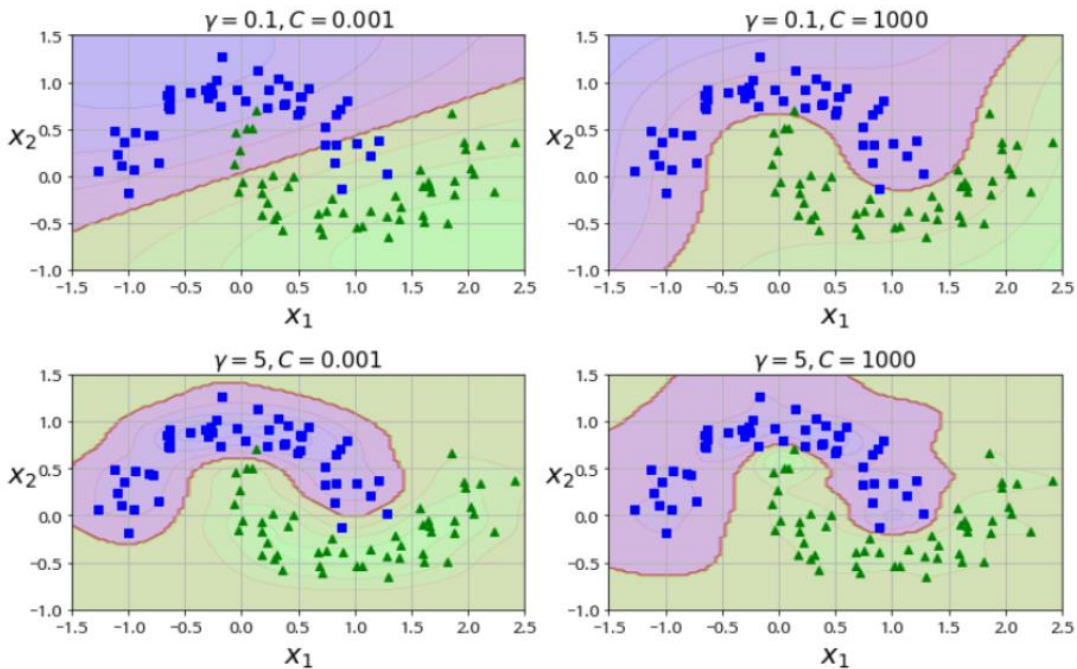
- 있음-> 하드마진
- 없음-> 소프트마진

데이터 샘플들이 다른 클래스에 놓이는걸 허용하는 정도

두 변수 모두 적정값을 찾아야 하며, 커질수록 복잡도가 증가한다.

일반적으로 grid search를 통해 최적 매개변수 값을 찾을 수 있다.

5.2 비선형 SVM 분류



※ 하이퍼파라미터인 γ 가 규제 역할 수행.
 - 모델이 과대적합? $\rightarrow \gamma$ 감소
 - 모델이 과소적합? $\rightarrow \gamma$ 증가

4) 계산 복잡도

Class	Time complexity	Out-of-core support	Scaling required	Kernel trick
LinearSVC	$O(m \times n)$	No	Yes	No
SGDClassifier	$O(m \times n)$	Yes	Yes	No
SVC	$O(m^2 \times n)$ to $O(m^3 \times n)$	No	Yes	Yes

• LinearSVC

- LinearSVM을 위한 최적화된 알고리즘을 구현한 liblinear 라이브러리 기반.
- 커널 트릭 지원x, 훈련 샘플과 특성 수에 거의 선형적으로 증가.
- 훈련시간 복잡도는 위와 같다.

• SVC

- 커널 트릭 알고리즘을 구현한 libsvm 라이브러리 기반.
- 시간 복잡도가 크기 때문에 훈련 샘플 수가 커지면 상당히 느려짐.
- 희소 특성이 몇 개 없는 경우에는 확장 원할.
- 알고리즘 성능은 샘플이 가진 희소 특성의 평균 수에 비례.

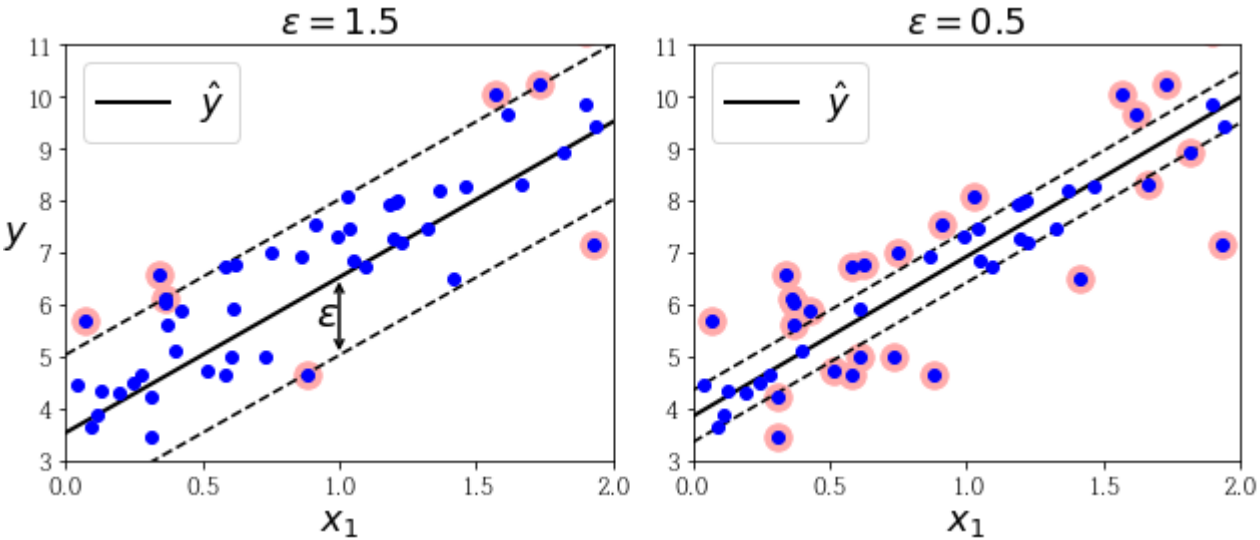
* 희소특성 (좀 더 자세히 알아보자)

: sparse features 대부분의 값이 0이거나 비어있는 특성 벡터. (<-> 밀집특성)

5.3 SVM 회귀

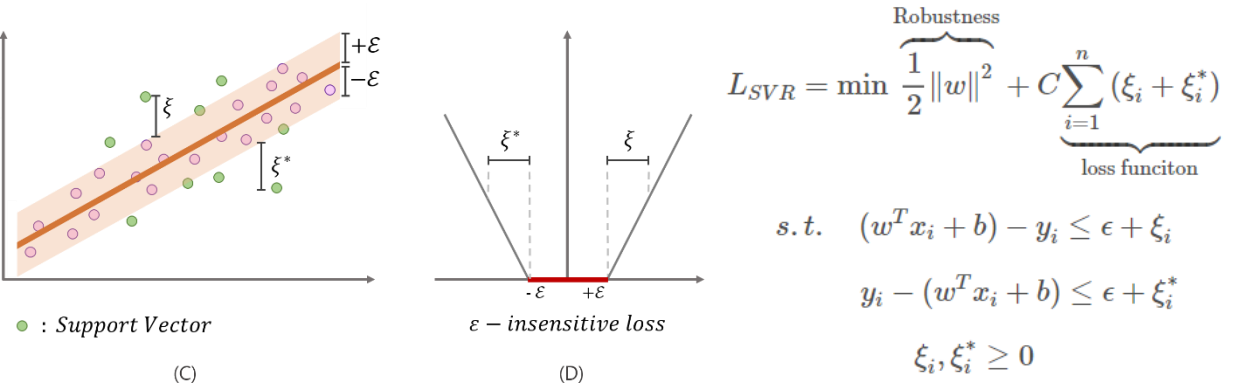
SVM을 회귀에 쓰고 싶다면, 제한된 마진 오류 안에서 가능한 한 많은 샘플이 들어갈 수 있게 학습하는 것이다. 해당 마진은 하이퍼파라미터(epsilon)로 조절할 수 있다!

<선형 모델 구현>



- 마진이 큰 왼쪽, 마진이 작은 오른쪽 (다만 예측에 큰 차이가 없다)
-> epsilon-intensive 하다. (엡실론에 민감하지 않다!)

<epsilon-insensitive 함수 사용한 SVR 식으로 표현한 SVR 손실함수>

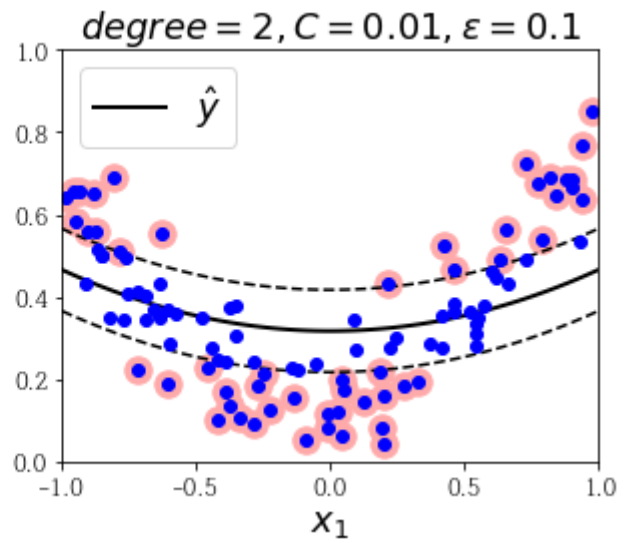
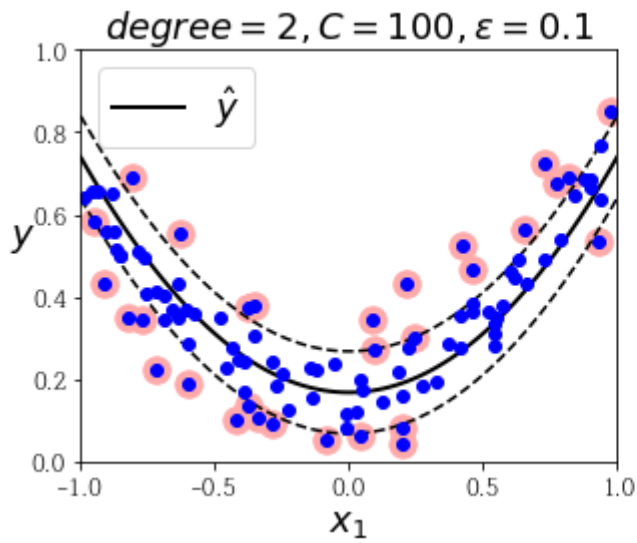


- ϵ : 회귀식 위아래 사용자가 지정한 값 \propto 허용하는 노이즈 정도
- ξ : 튜브 밖에 벗어난 거리 (회귀식 위쪽)
- ξ^* : 튜브 밖에 벗어난 거리 (회귀식 아래쪽)
- 튜브 내에 실제값 = penalty 0
- 튜브 밖에 실제값 = penalty는 C의 배율

SVR은 데이터에 노이즈가 있다고 가정하며, 이러한 점을 고려하여 노이즈가 있는 실제 값을 완벽히 추정하는 것을 추구하지 않는다. 따라서 적정 범위(2ϵ) 내에서는 실제값과 예측값의 차이를 허용한다.

5.3 SVM 회귀

<비선형 모델 구현>



- 규제가 거의 없는 왼쪽, 규제가 많은 오른쪽
-> 규제가 많을수록 그래프 커브는 가파르게 변한다.