

SVM

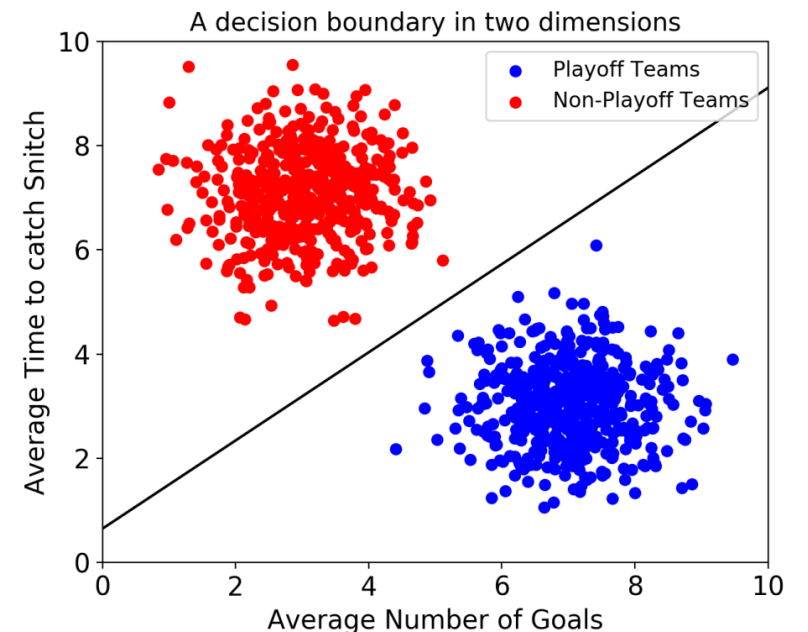
(Support Vector Machine)

16기 김지수

Support Vector Machine

■ Support Vector Machine

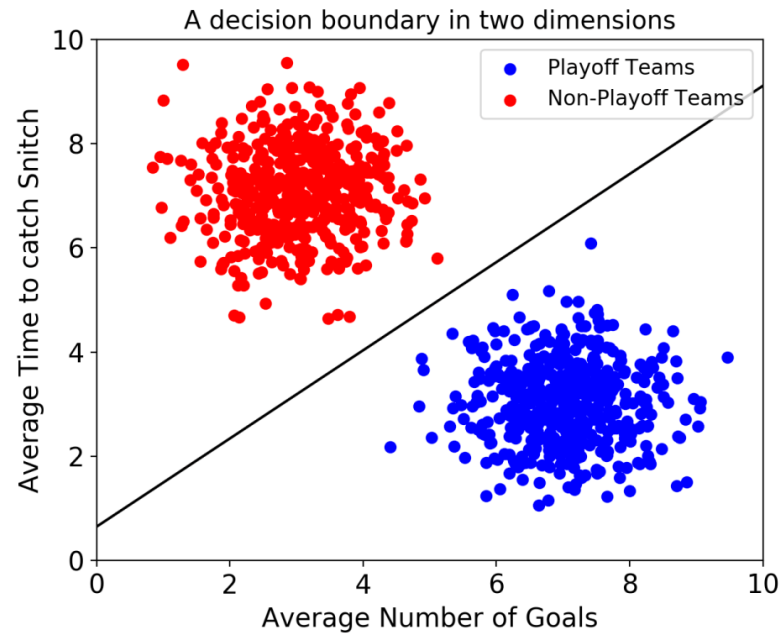
- » 기계 학습의 분야 중 하나로 패턴 인식, 자료 분석을 위한 지도 학습 모델
- » **결정 경계**, 즉 분류를 위한 기준 선을 정의하고 회귀분석에 주로 사용
- » 분류되지 않은 새로운 점이 나타나면 경계의 어느 쪽에 속하는지 확인해서 분류 과제를 수행
- » SVM 알고리즘은 그 중 가장 큰 폭을 가진 경계를 찾는 알고리즘



Support Vector Machine

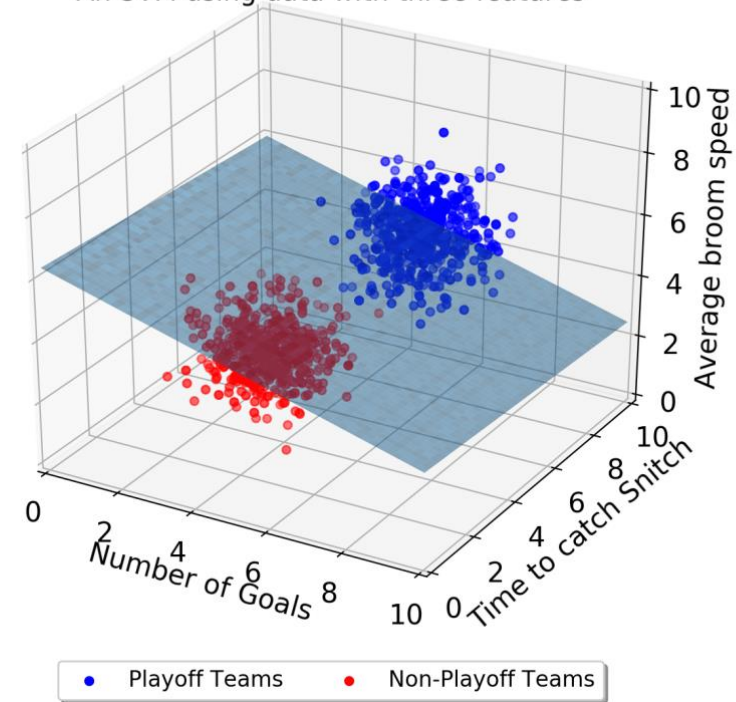
■ 결정경계 (Decision Boundary)

속성이 2개일 때



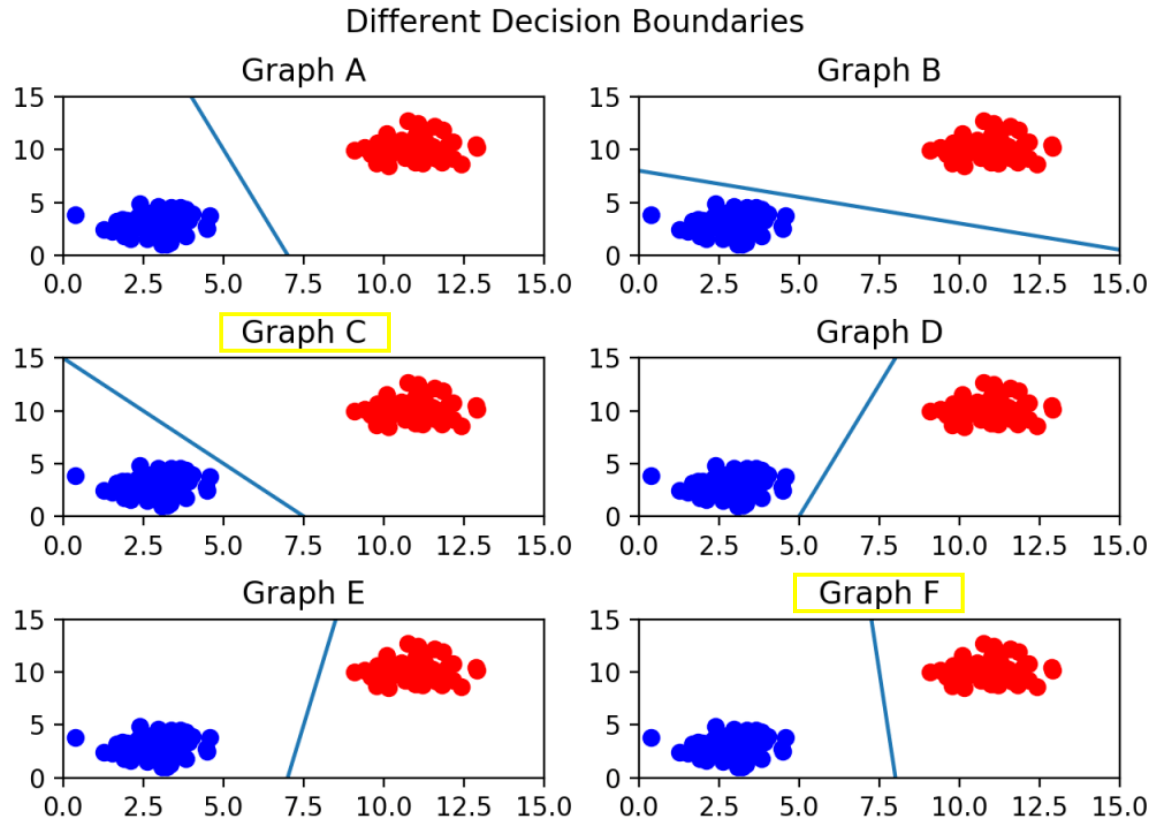
속성이 3개일 때

An SVM using data with three features



Support Vector Machine

■ 최적의 결정 경계



» Graph C -> 선이 파란색 부류와 너무 가까움

» Graph F -> 두 클래스 사이에서 거리가 가장 멀

» 결정 경계는 데이터 군으로부터 최대한 멀리 떨어지는 게 좋다

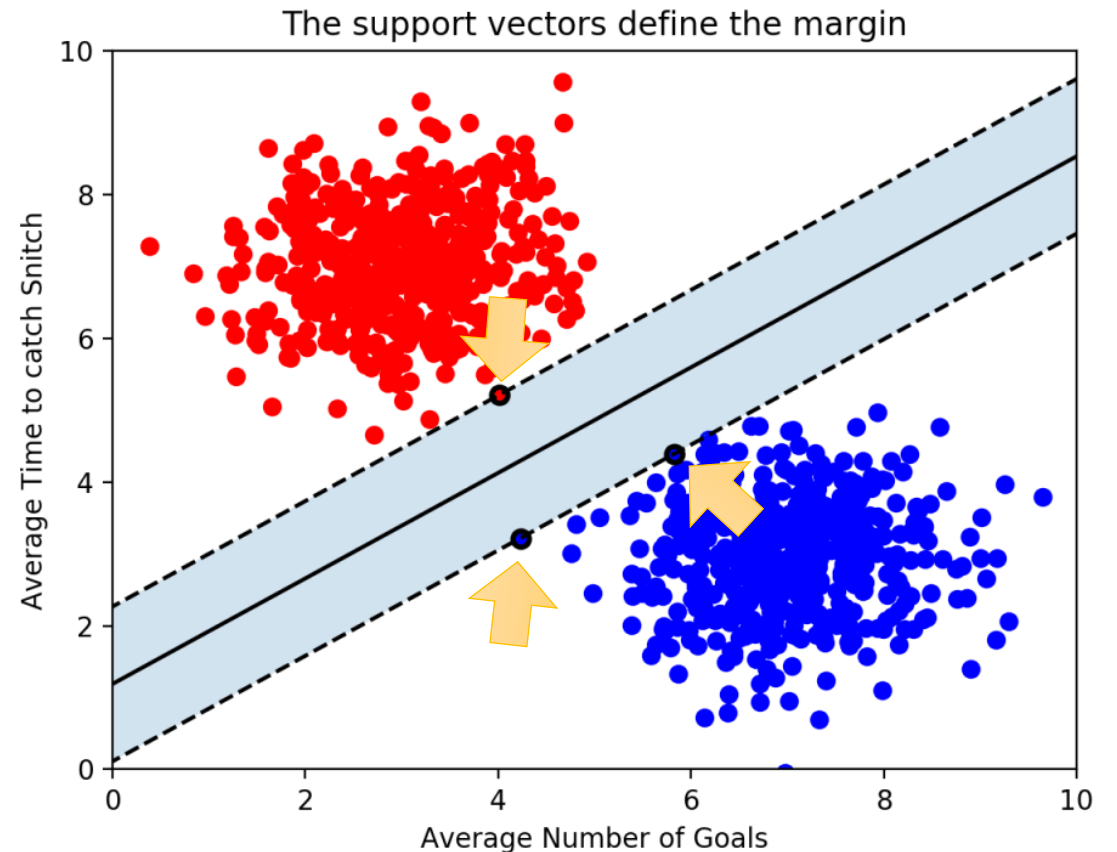
Support Vector Machine

Support Vectors

- » 결정 경계와 가까이 있는 데이터 포인트들을 의미
- » 경계를 정의하는 결정적인 역할

마진 (Margin)

- » 결정 경계와 서포트 벡터 사이의 거리
- » 최적의 결정 경계는 마진을 최대화한다.
- » n 개의 속성을 가진 데이터에는
최소 $n+1$ 개의 서포트 벡터가 존재한다



Support Vector Machine

■ SVM의 장점

- » 대부분의 머신러닝 지도 학습 알고리즘은 학습 데이터 모두를 사용하여 모델을 학습
- » SVM에서는 서포트 벡터만 잘 골라내면 나머지 쓸 데 없는 수많은 데이터 포인트들을 무시 가능
- » **속도가 매우 빠르다.**

Support Vector Machine

■ scikit-learn 사용법

```
from sklearn.svm import SVC
classifier = SVC(kernel = 'linear')
training_points = [[1, 2], [1, 5], [2, 2], [7, 5], [9, 4], [8, 2]]
labels = [1, 1, 1, 0, 0, 0]
classifier.fit(training_points, labels)
```

» 빨간색: 1, 파란색: 0

```
print(classifier.predict([[3, 2]]))
```

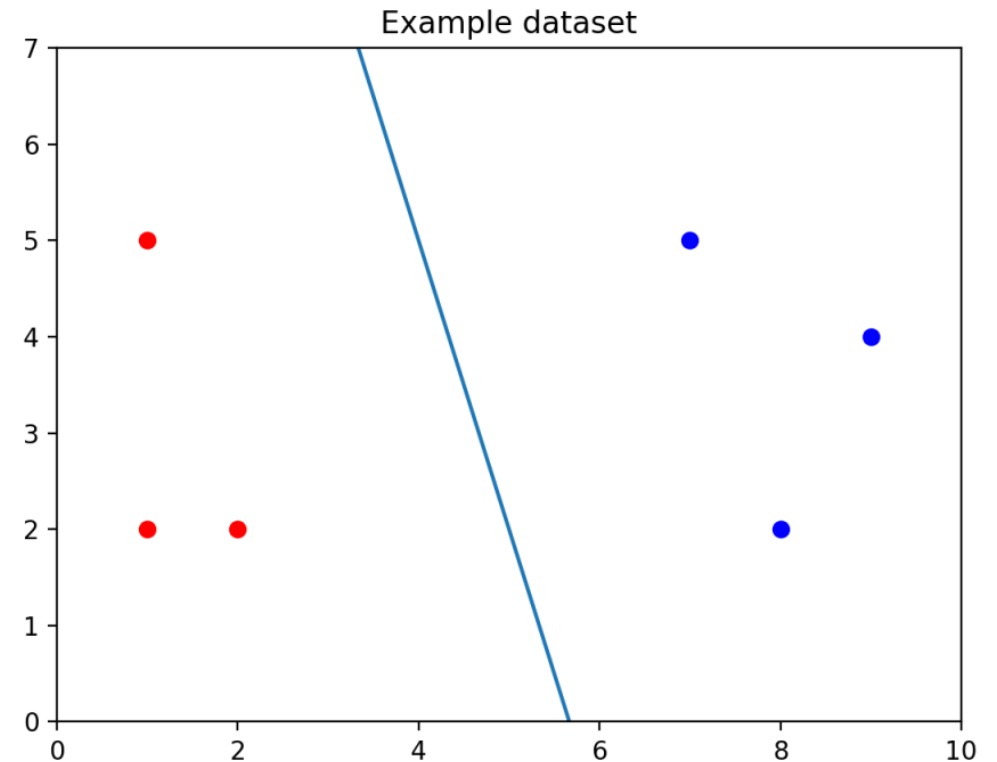
» 1로 분류

```
classifier.support_vectors_
```

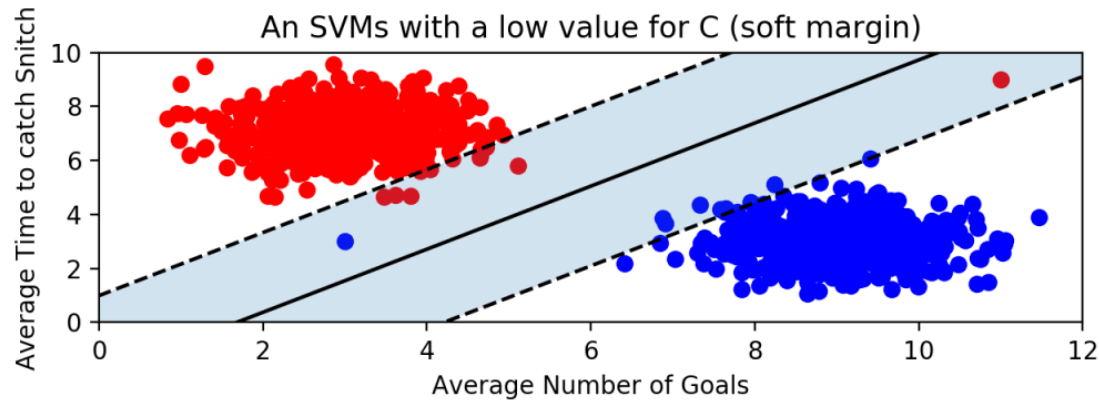
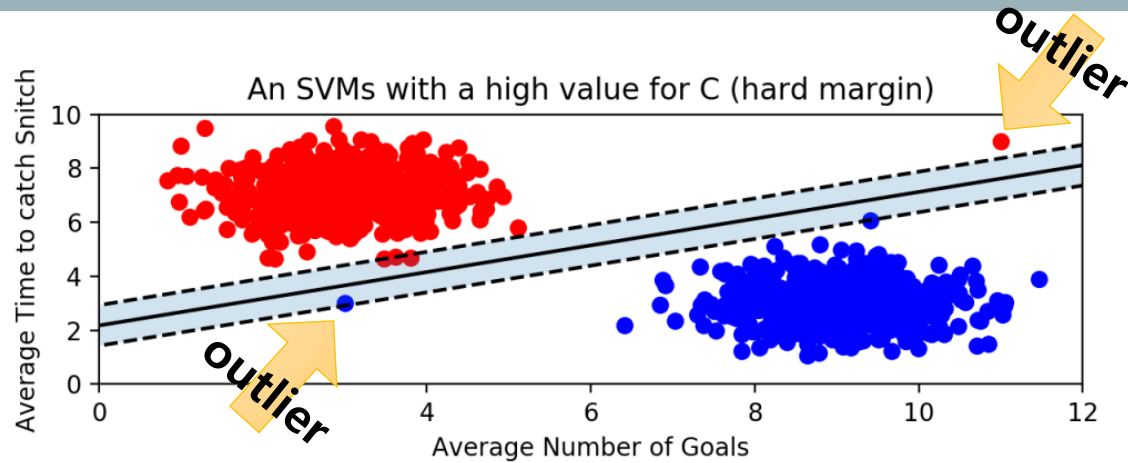
```
[[7, 5],
```

```
[8, 2],
```

```
[2, 2]]
```



Support Vector Machine



하드 마진(hard margin)

- » 서포트 벡터와 결정 경계 사이의 거리가 매우 좁다
- » 마진이 작아진다
- » 오버피팅 문제 발생

소프트 마진(soft margin)

- » 서포트 벡터와 결정 경계 사이의 거리가 멀어진다
- » 마진이 커진다
- » 언더피팅 문제 발생

파라미터 C (기본값=1)

```
classifier = SVC(C = 0.01)
```

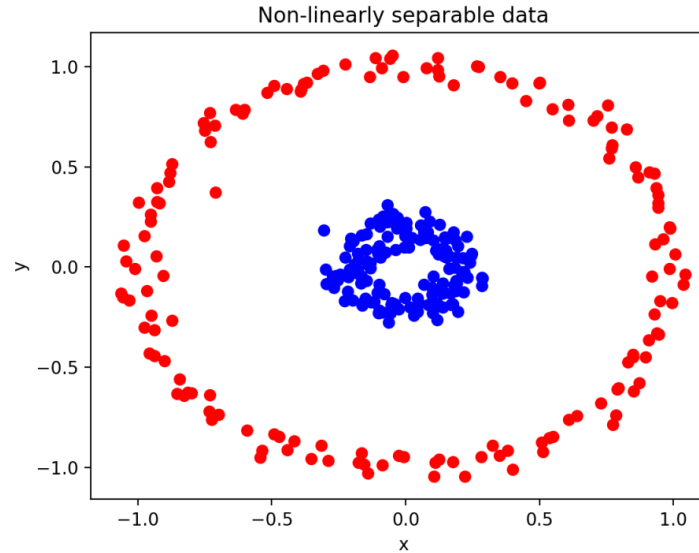
C의 값이 클수록 하드마진(오류 허용 안 함), 작을수록 소프트 마진(오류 허용)

Support Vector Machine

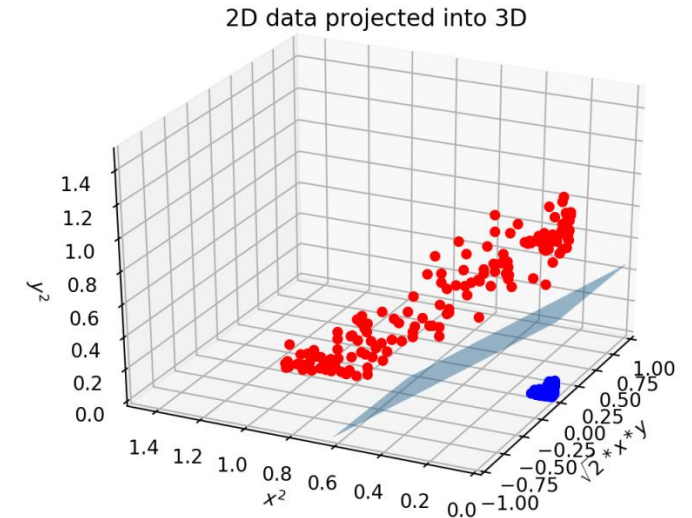
■ 커널(Kernel) (기본값=rbf)

■ 다항식 (Polynomial)

```
from sklearn.svm import SVC
classifier = SVC(kernel='poly')
```



2차원



3차원

■ 방사 기저 함수 (RBF: Radial Bias Function)

```
from sklearn.svm import SVC
classifier = SVC(kernel='rbf')
```

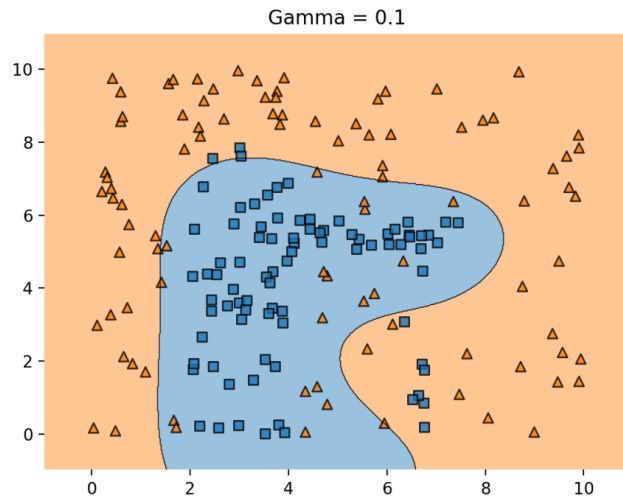
2차원 → 무한 차원

Support Vector Machine

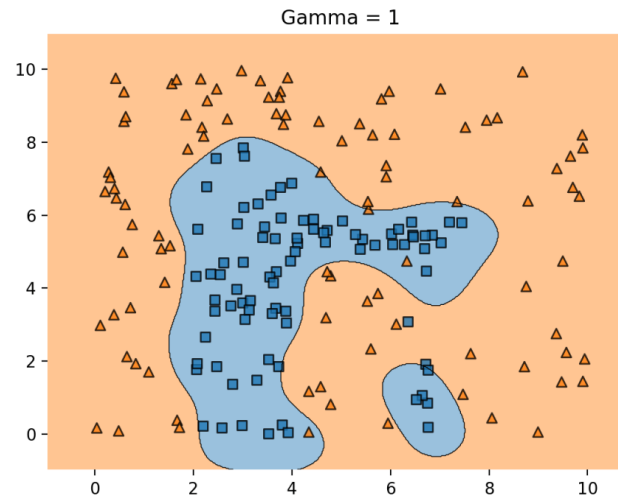
■ 파라미터 gamma

» 결정 경계를 얼마나 유연하게 그을 것인지 정해줌

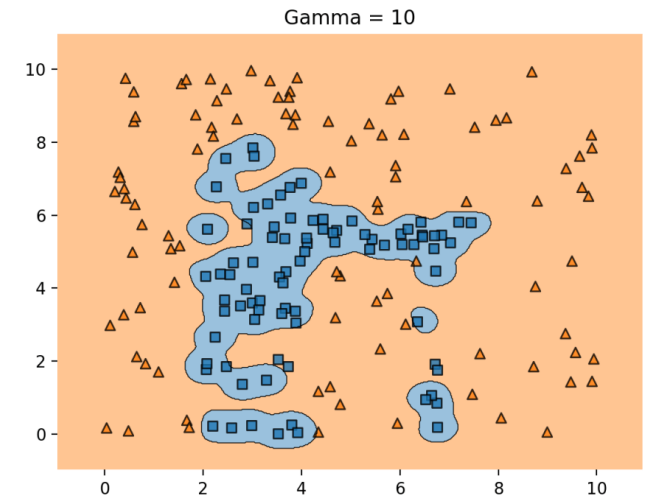
```
classifier = SVC(kernel = "rbf", C = 2, gamma = 0.5)
```



» 학습데이터에 별로 의존 안함
» 언더피팅



» 적당



» 학습데이터에 많이 의존
» 오버피팅

Thank you