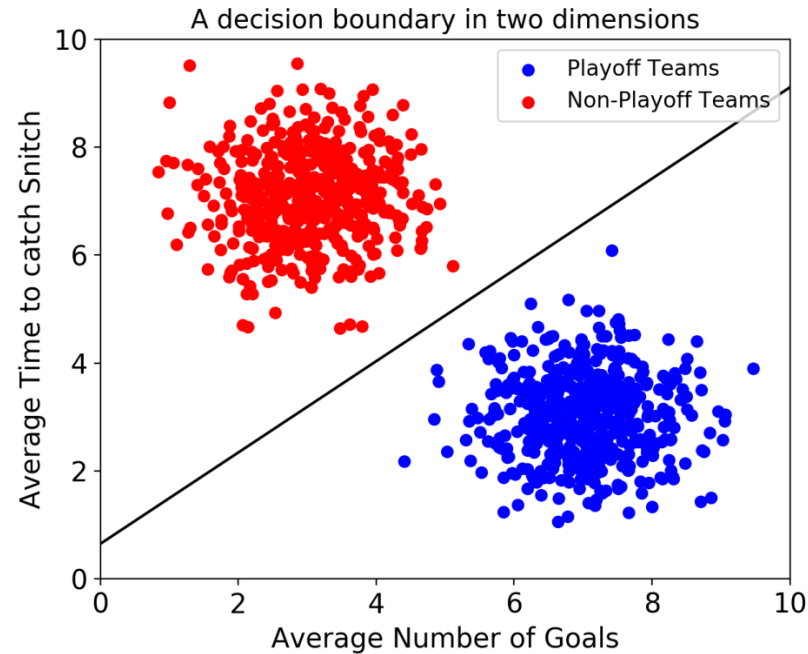


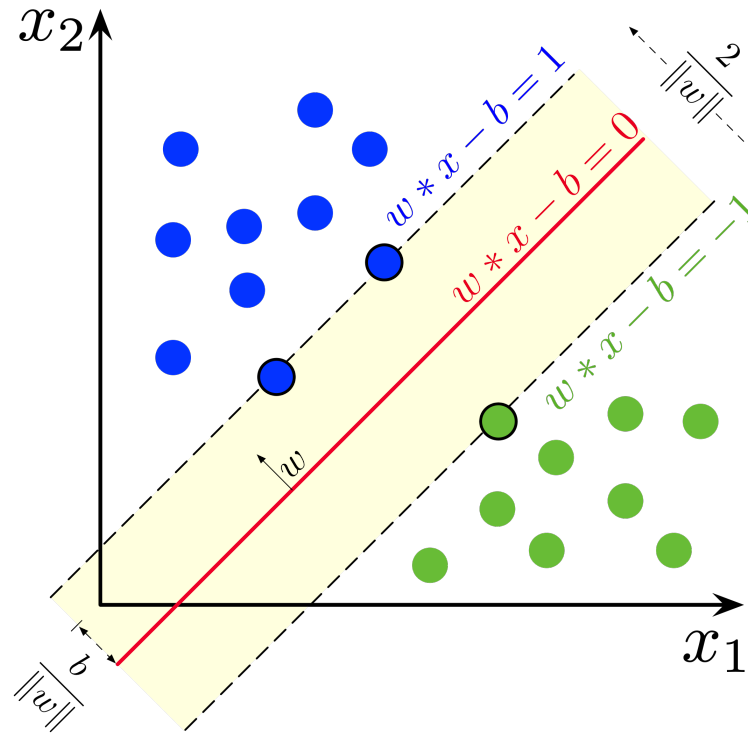


Support Vector Machine (SVM)



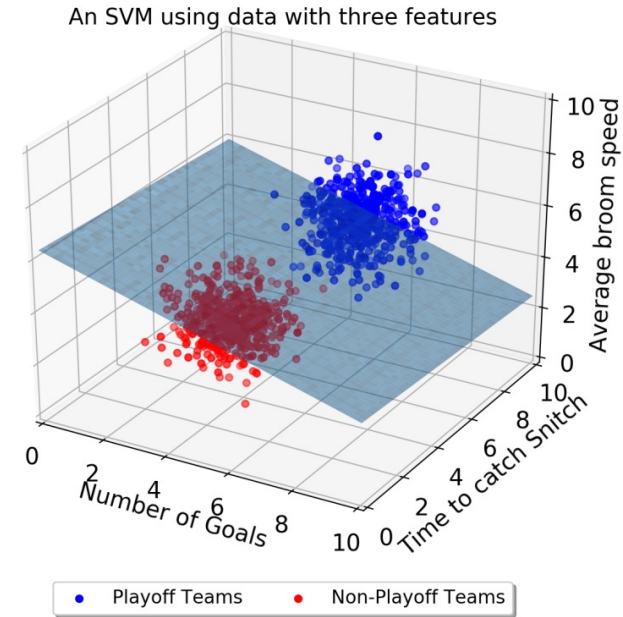
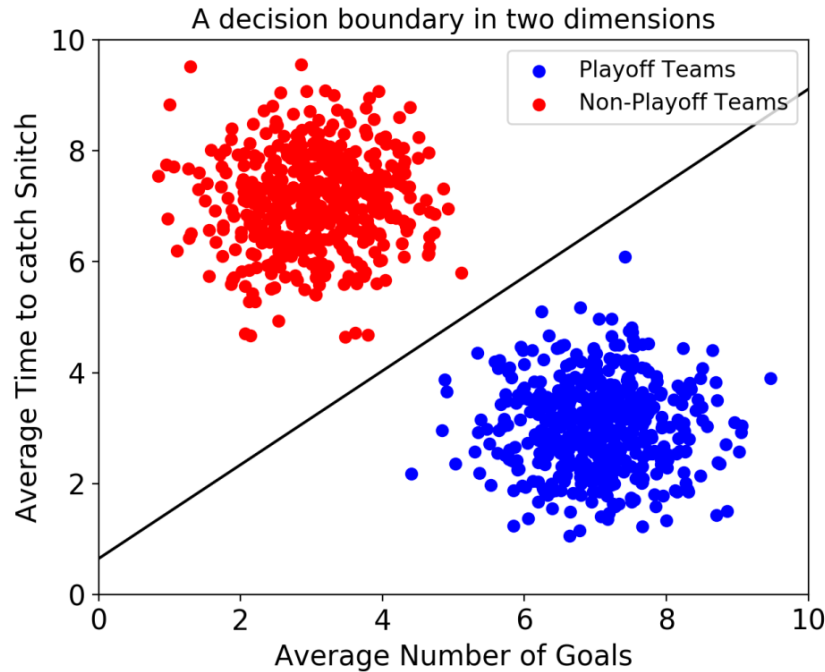
- 결정경계(decision boundary)를 이용한 classification, regression, clustering 모델
- 최초 제안된 모델은 classification 모델이었으나, 이후 다른 분석도 가능
- Logistic, Naïve Bayes와 달리 ‘예측확률’을 구할 수 없다

결정경계 (Decision Boundary)



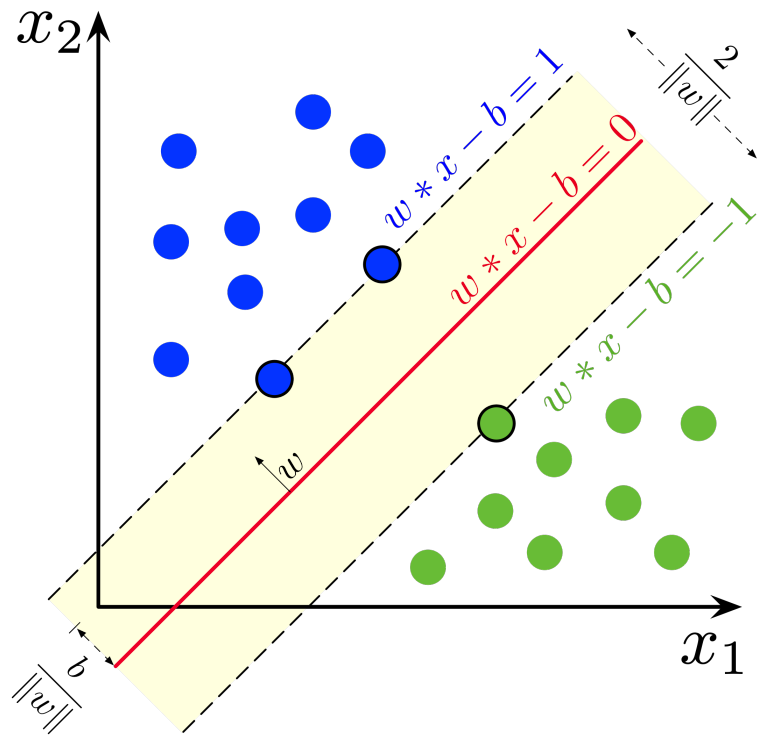
- 선에 가장 가까운 양쪽 점을 수직이등분하는 선인듯?
• Support vector를 지나는 초평면의 거리를 극대화하는 w 를 계수로 하는 초평면을 decision boundary로 사용

초평면(Hyper plane)



- 결정 경계는 feature의 개수에 따라서 차원이 달라진다.
(2개의 feature의 경우 → 1차원의 선, 3개의 feature의 경우 → 2차원의 평면)
- feature의 개수가 늘어날 수록 경계는 고차원의 형태를 띄는데, 이를 초평면이라고 부름

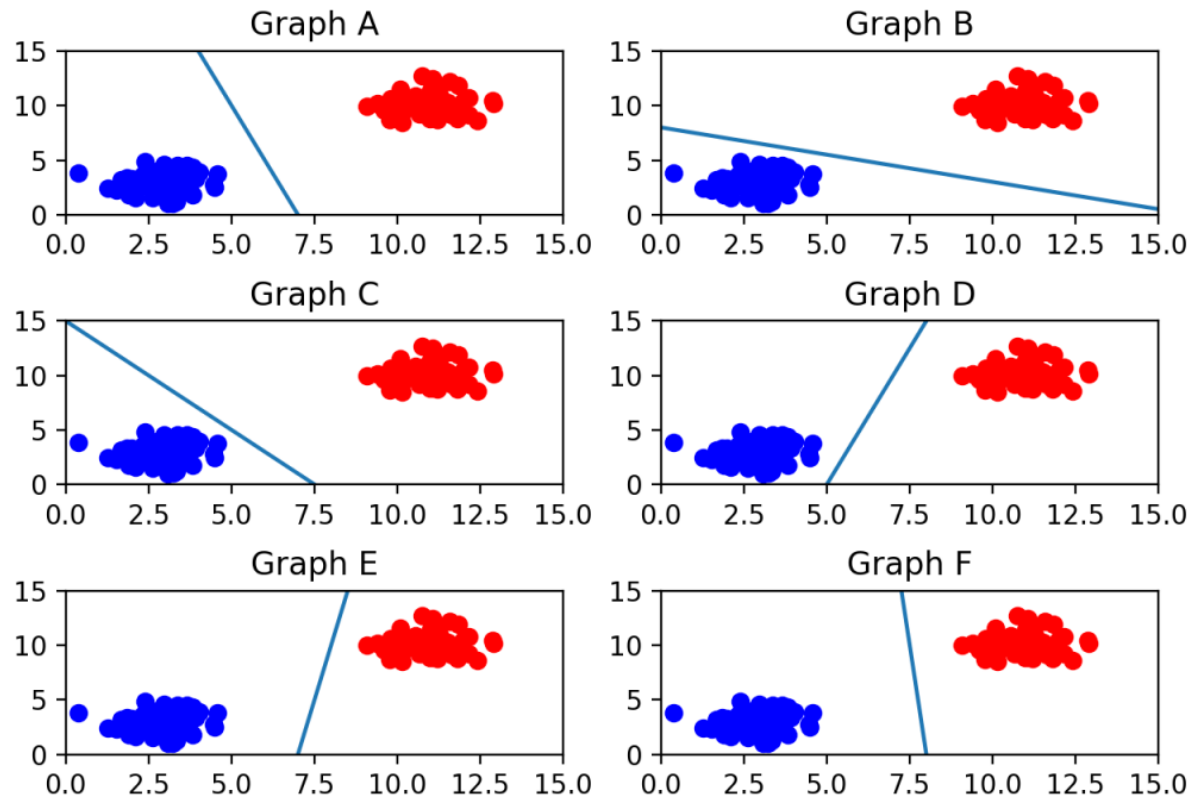
서포트 벡터 (Support Vector)



- 서포트 벡터는 결정 경계 위의 관찰 값들을 의미한다.
- 서포트 벡터는 결정 경계를 결정하는 표본들이다.
- 서포트 벡터를 움직이면 길의 폭이 달라진다.
- 서포트 벡터 이외의 관찰 값들은 길의 폭을 결정하는 데에 아무런 영향을 미치지 못한다.

최적의 결정경계

Different Decision Boundaries



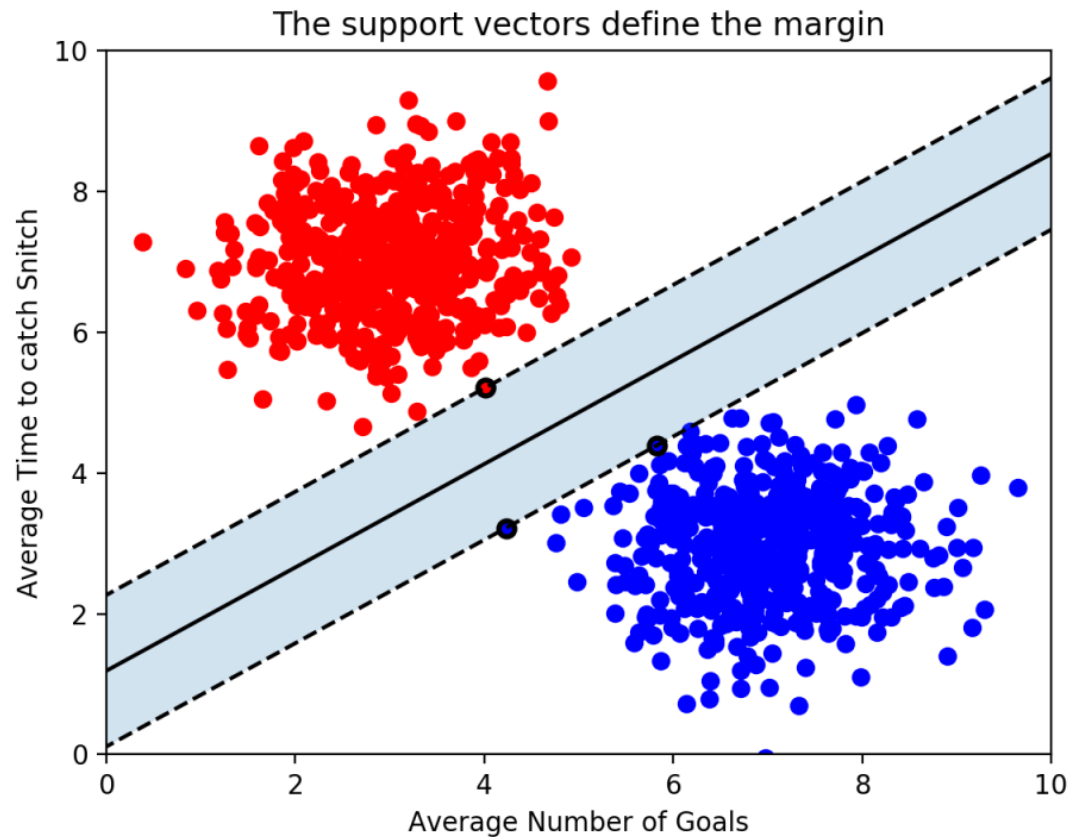
1. 어떤 그래프가 가장 위태로워 보이는가? **거리가 제일 가까운거**

2. 그렇다면 어떤 결정경계가 가장 적절해 보이는가? **거리가 제일 먼거**

• 결정경계는 데이터 군으로부터 최대한 **멀리** 떨어지는 것이 좋다!! Why??

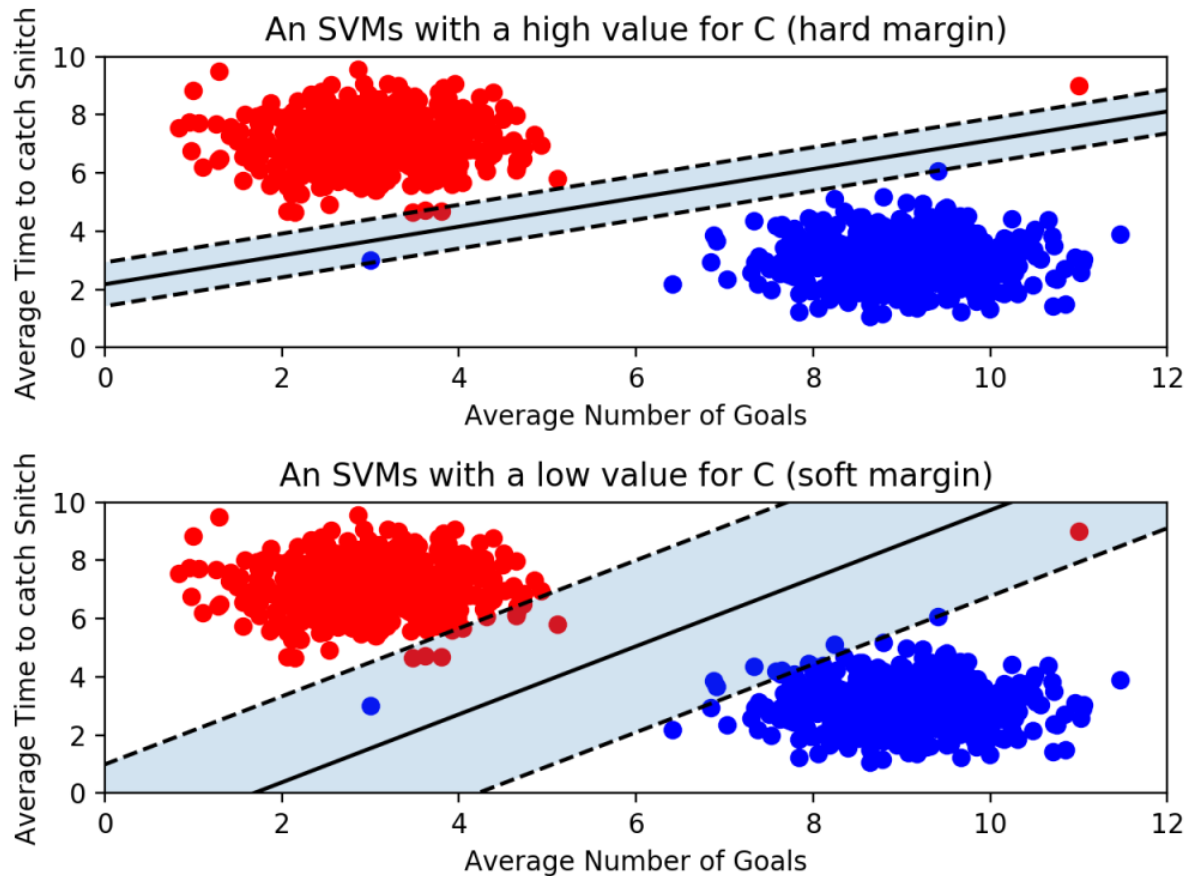
이 거리를 마진이라고 함
최적의 결정경계는 마진을 최대화함

마진 (Margin)



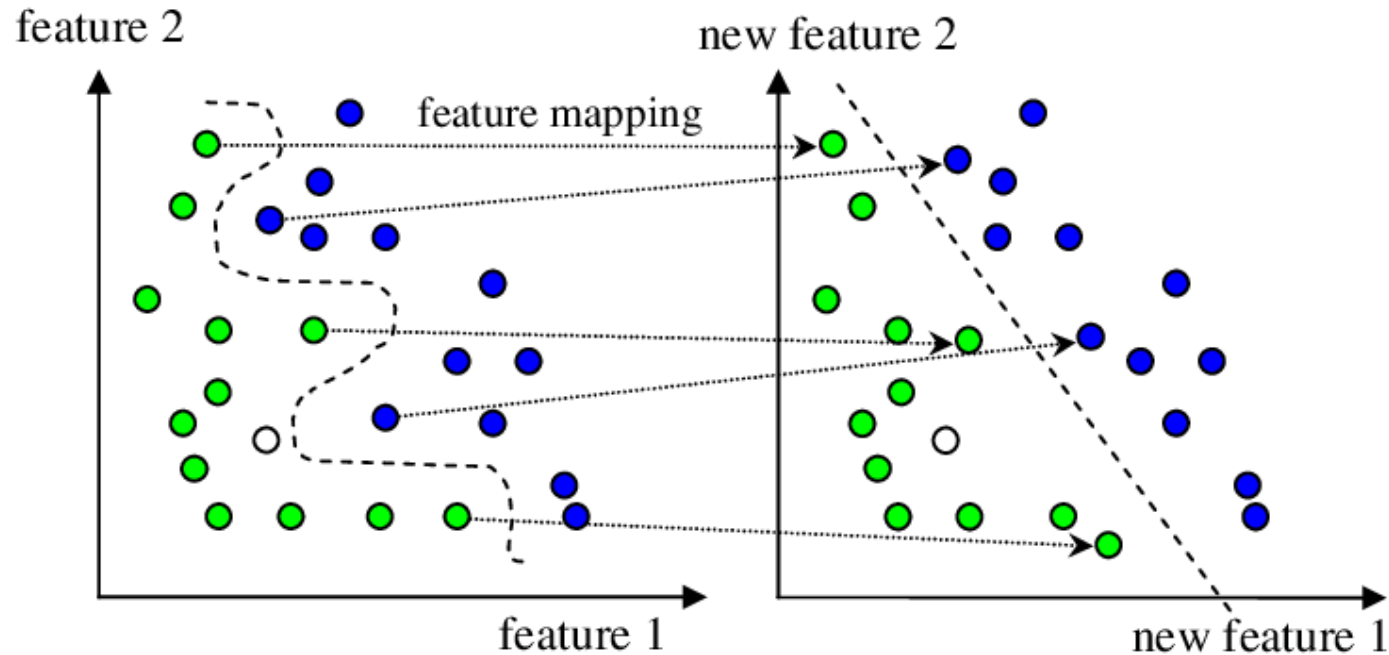
- Support vector로부터 결정 경계 사이의 거리
- 최적의 결정 경계는 마진을 최대화한다.
- n 개의 속성을 가진 데이터에는 최소 $n+1$ 개의 서포트 벡터가 존재한다.

Hard Margin vs Soft Margin



- **Hard Margin**
서포트 벡터와 결정 경계의 사이가 매우 좁다.
outlier들을 허용하지 않는다.
오버피팅의 문제가 발생할 수 있다.
- **Soft Margin**
서포트 벡터와 결정 경계의 사이가 매우 넓다.
outlier들을 허용한다.
언더피팅의 문제가 발생할 수 있다.
- 파라미터 C의 값이 클 수록 마진이 작아진다.
최적의 C값은 모델을 계속 수정하면서 찾아내야 한다.

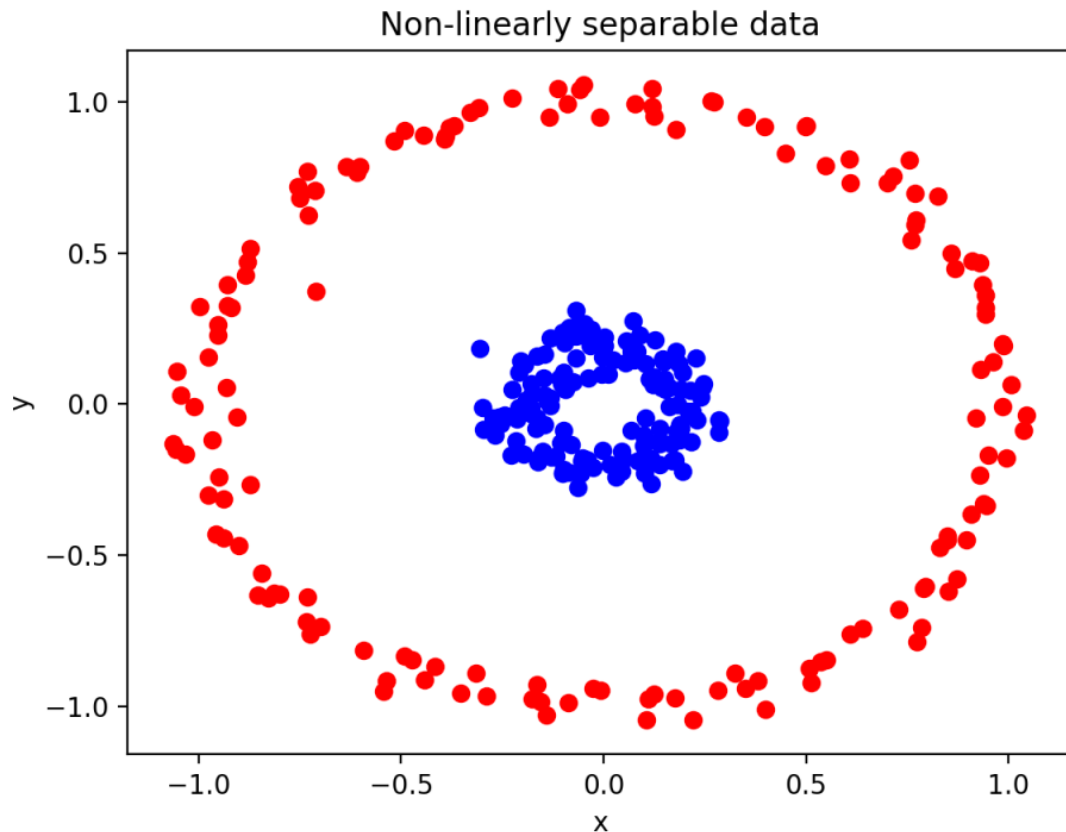
커널 (Kernel)



- 차원 확장과 비선형 decision boundary
눈에 띄는 차이가 있지만 선형으로 분류하기 '어려운' 자료라면?
→ feature를 비선형으로(차원확장) or decision boundary를 비선형으로

비선형 decision boundary는 함수의 형태를 특정하기가 쉽지 않다.
다만 자료의 비선형변환은 계산은 복잡해지지만 선형 함수를 유지하면서 안정적이다.

커널 (Kernel)



- 만약 SVM이 선형으로 분리할 수 없는 데이터 셋이 있다면 어떻게 해야할까?
>> kernel을 지정하여 해결할 수 있다.
- 단순히 outlier 때문에 선형으로 분리할 수 없다고 판단해서는 안된다
>> 일부 outlier들을 분류하기 위해, 비선형으로 결정 경계를 만들 필요가 없다
- 모든 점을 올바르게 분리하는 선을 그린다면, 모델에 데이터가 과도하게 적합되는 overfitting이 발생한다.

커널 (Kernel)

- Linear : 단순한 내적

$$k(x, z) = x \cdot z$$

- Polynomial : p차까지의 항을 포함

$$k(x, z) = (1 + x \cdot z)^p$$

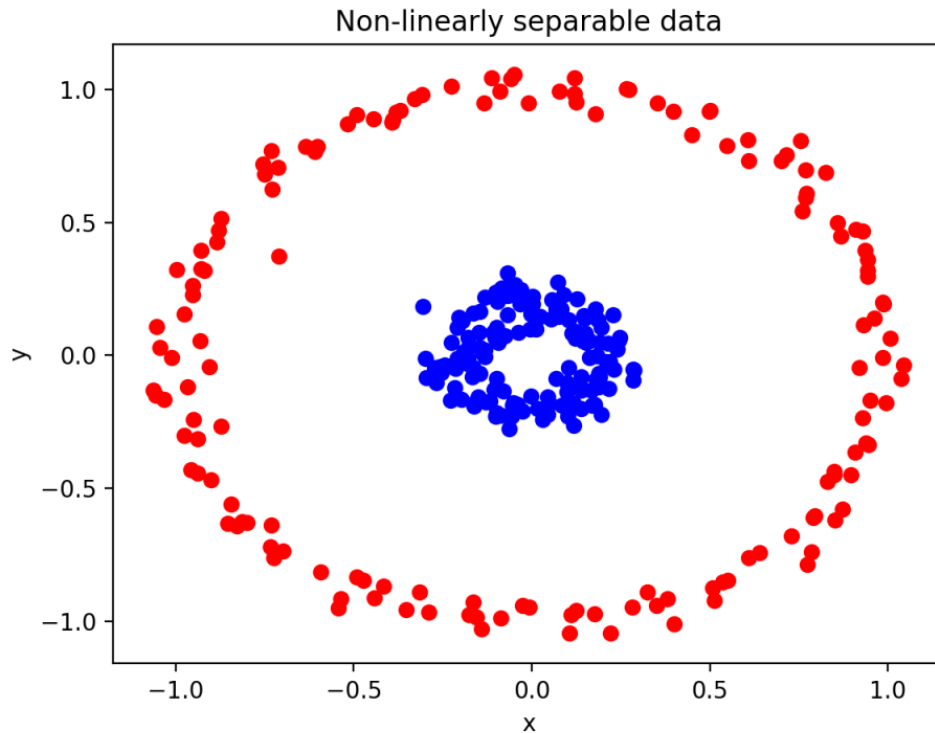
- Radial Basis Function (RBF) : 차수가 무한인 다항식

$$k(x, z) = \exp\left(-\frac{(x - z)^T(x - z)}{2\sigma^2}\right) \text{ for } \sigma > 0, \exp(\alpha x \cdot z) = \left(\frac{\sum_{i=0}^{\infty} \frac{\alpha^i (x \cdot z)^i}{i!}\right)$$

- Sigmoid : logistic과 유사

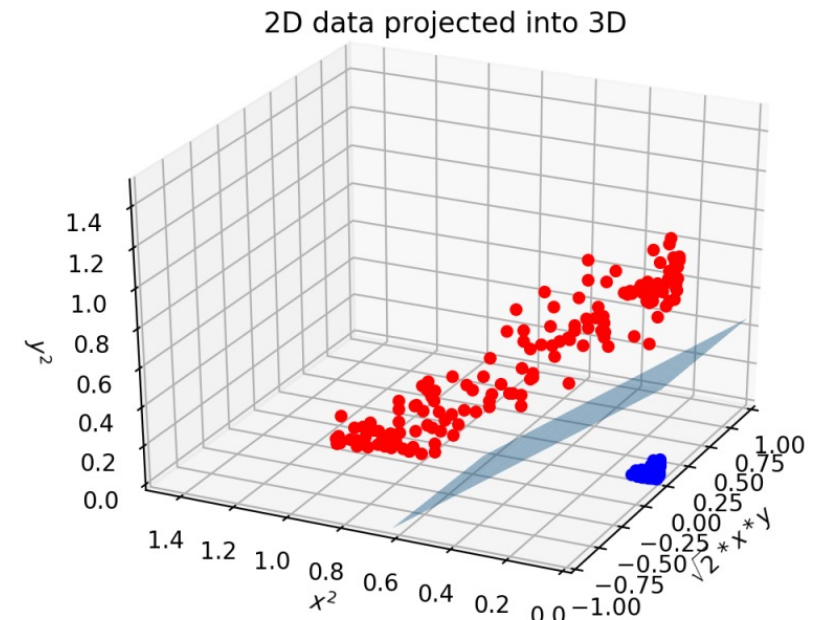
$$k(x, z) = \tanh(\beta_0 x^T z + \beta_1)$$

커널 (Kernel)

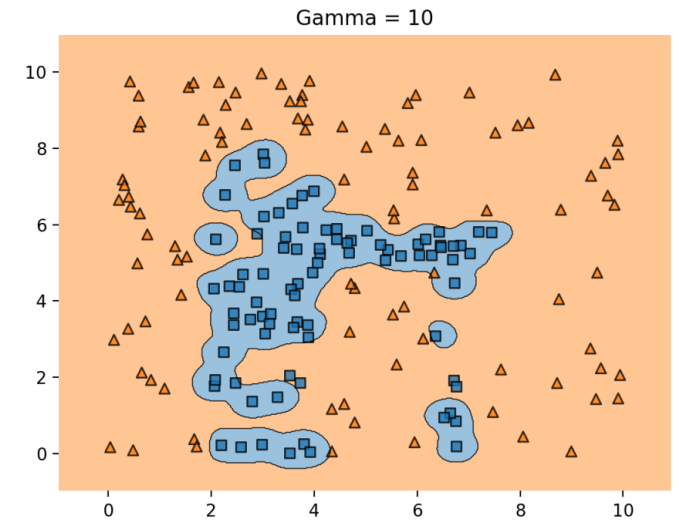
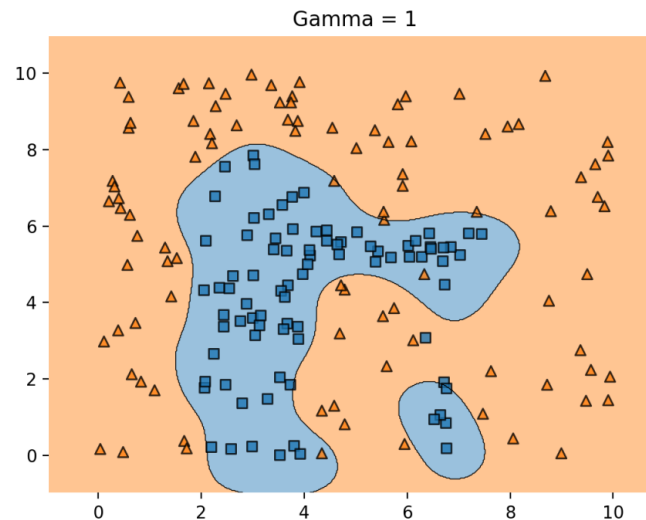
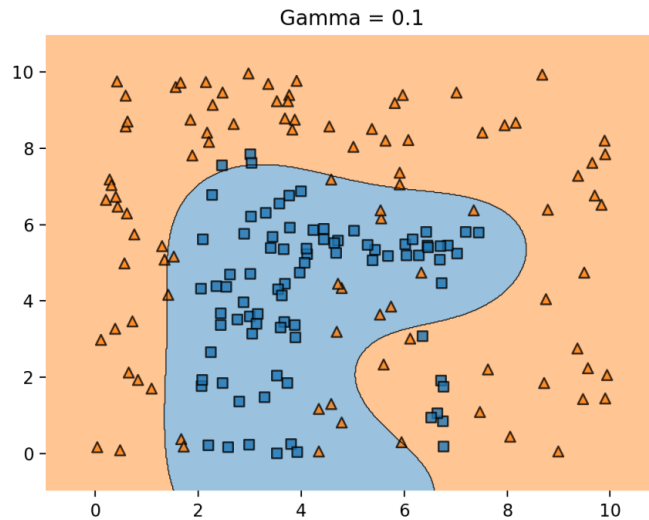


ex) kernel = poly

$$(x, y) \rightarrow (\sqrt{2} \cdot x \cdot y, x^2, y^2)$$
$$(1, 2) \rightarrow (2\sqrt{2}, 1, 4)$$



커널 (Kernel)



ex) kernel = rbf

RF 커널의 gamma는 결정 경계를 얼마나 유연하게 하는지를 결정하는 파라미터

gamma값을 높이면 학습 데이터에 많이 의존하여 결정 경계를 구불구불하게 만든다 → overfitting

gamma값을 낮추면 학습 데이터에 덜 의존하여 결정 경계를 직선에 가깝게 긋는다 → underfitting

* 결정경계가 비선형으로 보이지만 차원을 확장해보면 이는 실제로는 선형 결정경계이다!

SVM의 장단점

Grid search

- SVM의 장점
 - 탄탄한 이론에 근거한 모형으로, 일반화가 쉬움
 - Feature의 수가 많은 text나 image 분류에 효율적이다
 - 모형의 복잡도는 feature의 수 보다는, 표본 수가 더 크게 좌우한다
 - 다중공선성에서 자유롭다 릿지, 라소, 엘라스틱
 - 정규화를 통해 soft margin SVM의 overfitting 문제를 쉽게 통제할 수 있다
 - 작은 표본에서 다른 모형들에 비해 상대적으로 성능이 좋다
- SVM의 단점
 - 상대적으로 시간이 많이 걸리며, 대형 자료에서 특히 시간이 오래 걸린다
 - 예측에 대한 확률을 추정하지 않는다
 - class의 관찰 값들이 많이 중첩되어 있을 경우, 효율성이 떨어진다.

