서포트 벡터 머신

Support Vector Machine

2021.01.30

이윤정

01 Support Vector Machine

SVM 이란

: 지도학습 중 분류모델에 해당되며, 회귀(SVR) 및 이상치 탐색에도 사용된다.

SVM의 분류

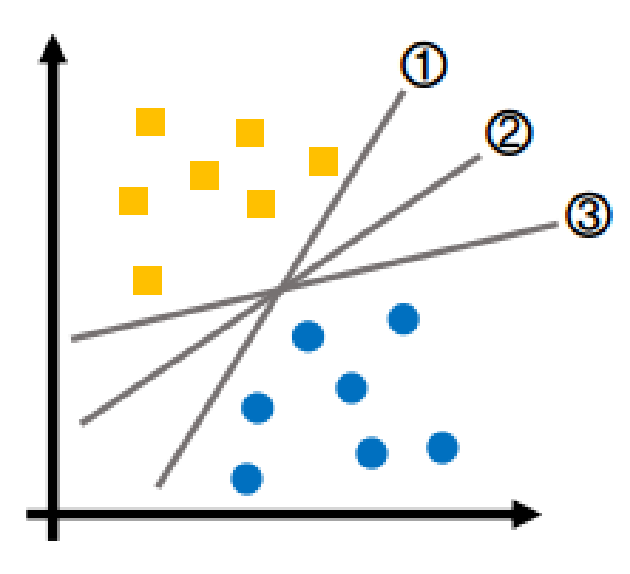
선형	분류
O	선형 SVM
X	비선형 SVM

error 허용 여부	분류
O	Soft margin SVM
X	Hard margin SVM

01 Support Vector Machine

SVM의 개념

: 주어진 데이터 점들이 2개의 그룹 안에 각각 속해 있다고,
가정 시 새로운 데이터 점이 두 그룹 중 어느 곳에 속하는 지 판단하는 알고리즘



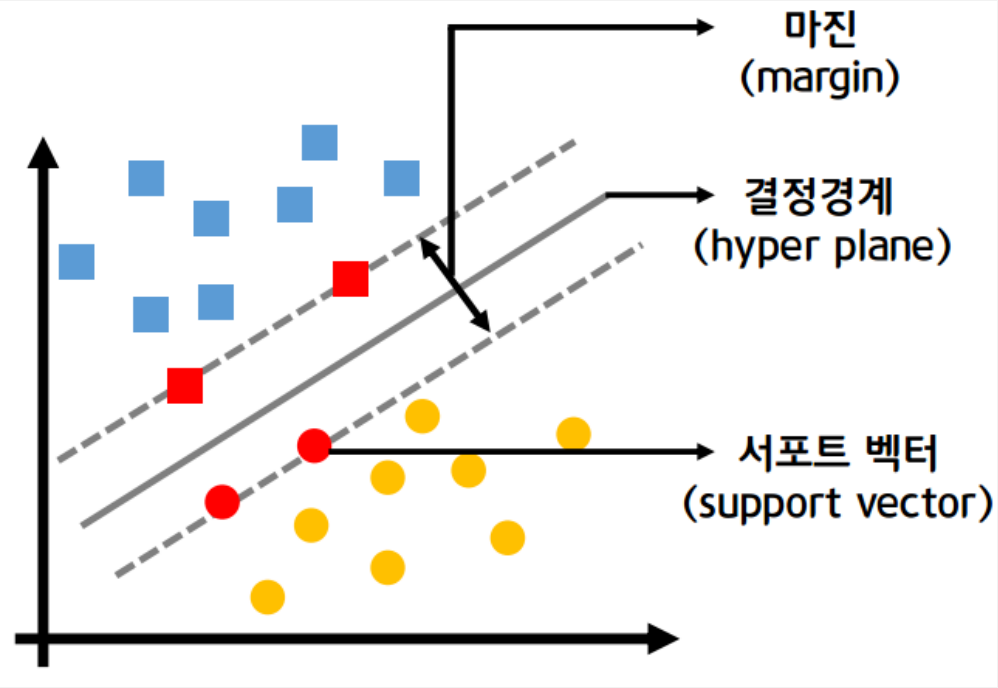
SVM의 관점에서 데이터를 가장 잘 분류한 선은 몇번일까요?

바로, 2번입니다.

원래 존재한 데이터와 경계선이 충분히 떨어져 있어 쉽게 구분 가능하기 때문
즉, 결정 경계(2번)과 데이터 간의 거리(여백)이 크다.

01 Support Vector Machine

SVM의 용어



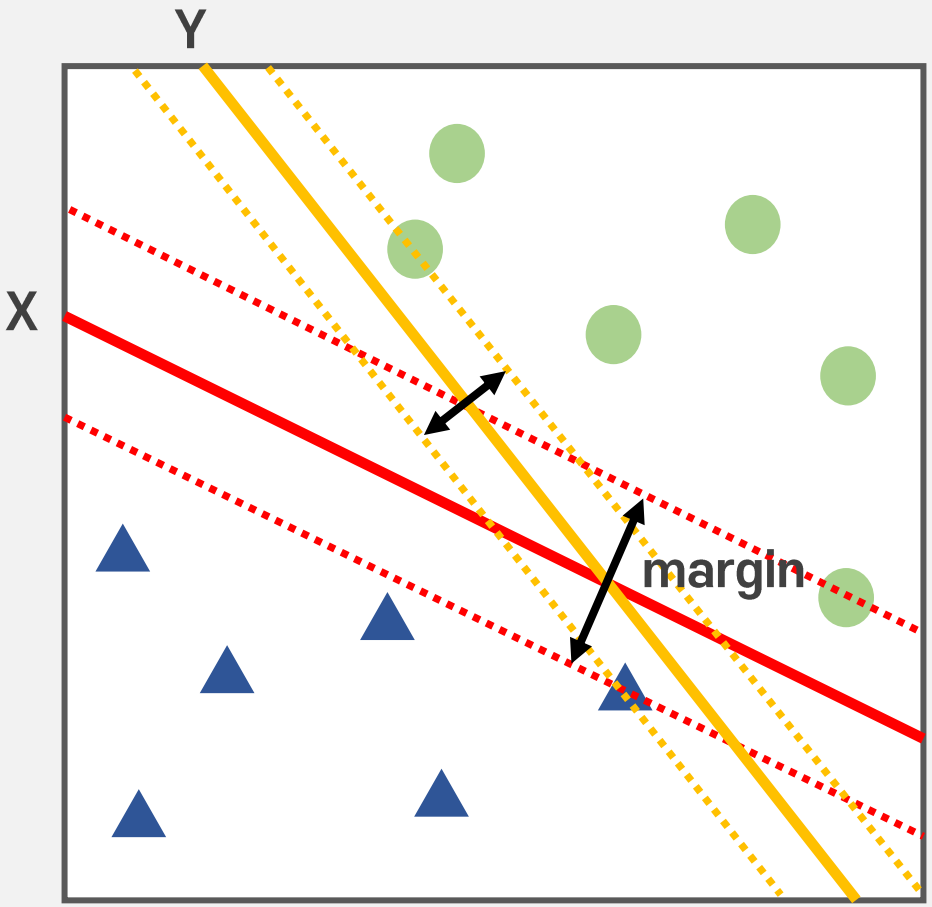
① Decision Boundary(결정 경계) = Hyperplane(초평면)
: 데이터를 나누는 기준이 되는 경계

② Support Vector : Hyperplane과 가장 가까운 data

③ Margin : 결정경계와 서포트벡터 사이의 거리 X 2

SVM의 알고리즘

: Margin을 최대화하는 결정 초평면(decision hyperplane)을 찾는 것



$WX + b = 1$

$WX + b = 0$

$WX + b = -1$

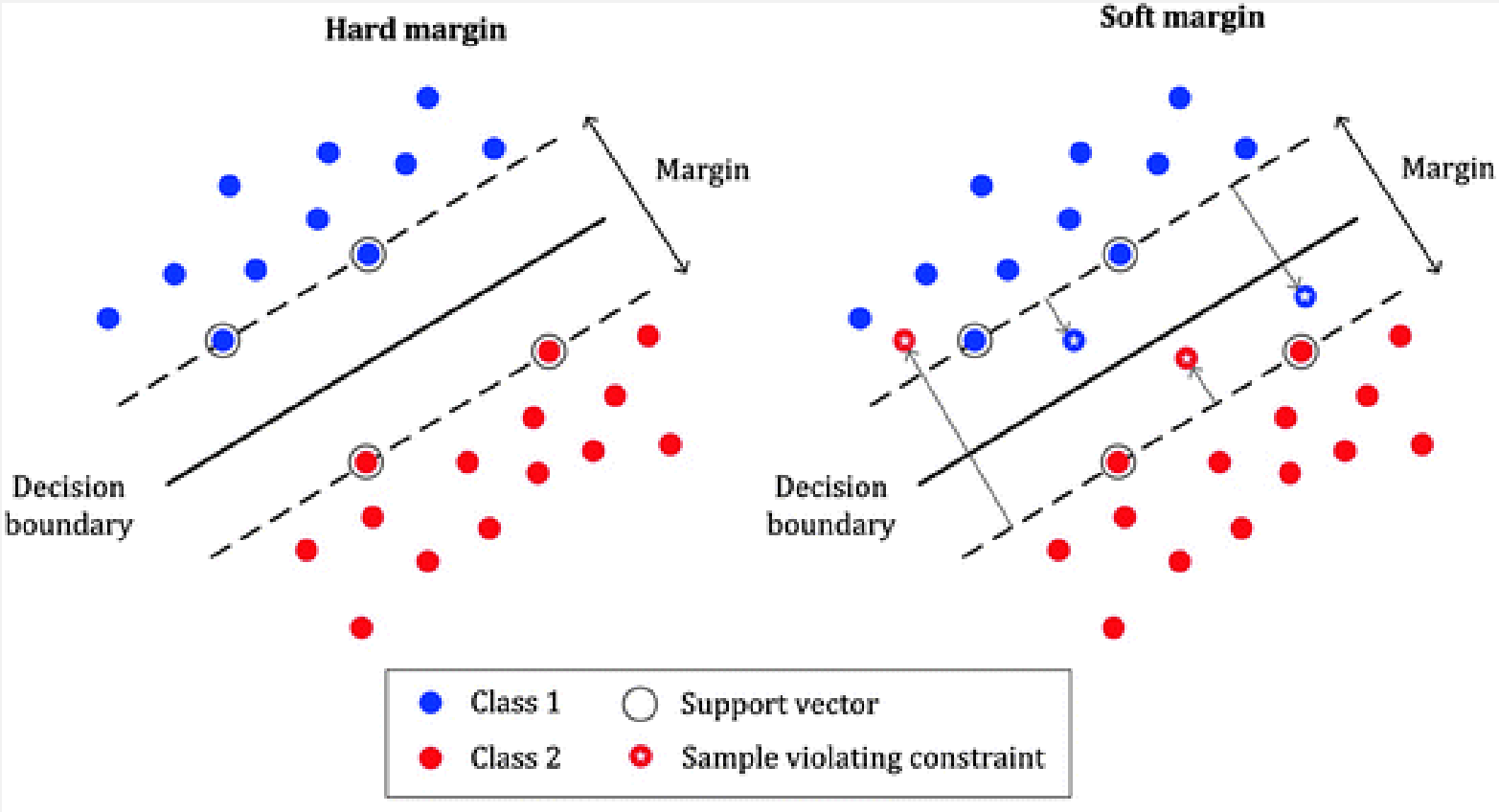
X와 Y 모두 결정초평면이되,
데이터 간의 margin을 최대화하는 결정초평면은 X이다.

$$\begin{aligned}\textbf{margin} &= \text{distance}(x^+, x^-) \\ &= \|x^+ - x^-\|^2 \\ &= (x^+ - x^-)\frac{w}{\|w\|} \\ &= \frac{1 - b - (-1 - b)}{\|W\|} \\ &= \frac{2}{\|W\|}\end{aligned}$$

02 Soft Margin SVM

Soft Margin SVM 이란

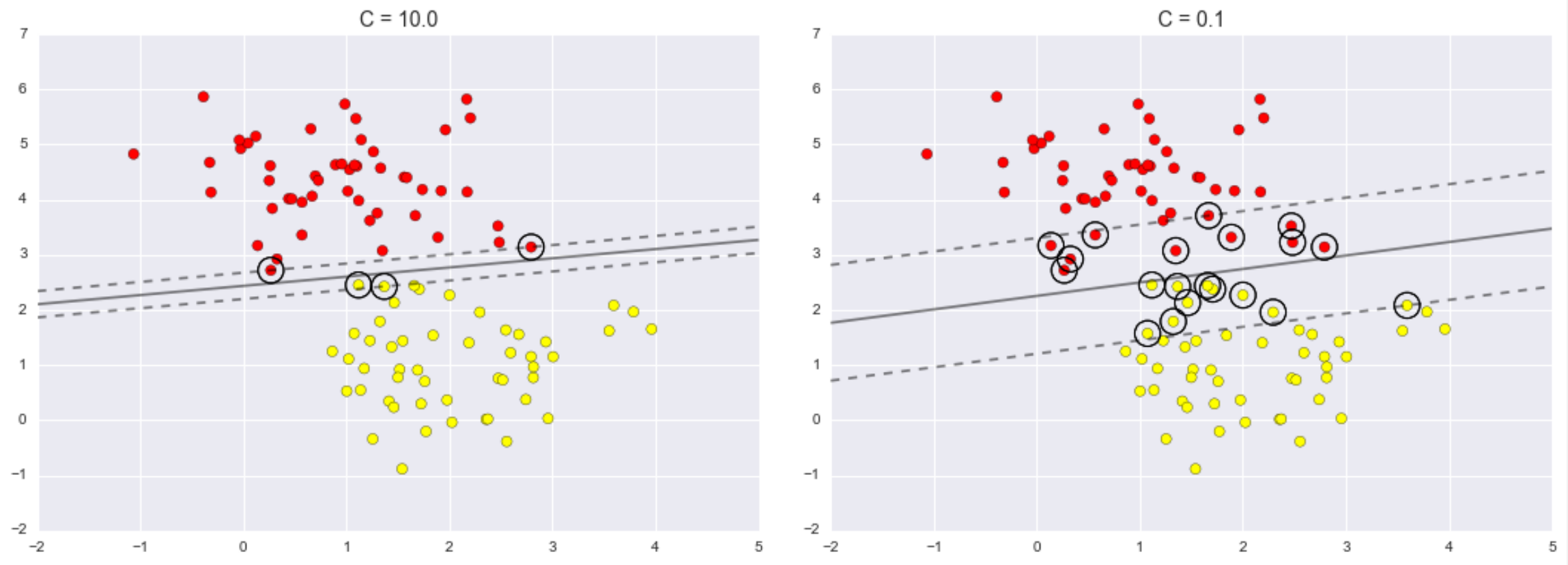
: error를 허용하되, 패널티를 통해 전체 error를 최소화



02 Soft Margin SVM

하이퍼 파라매터 C란

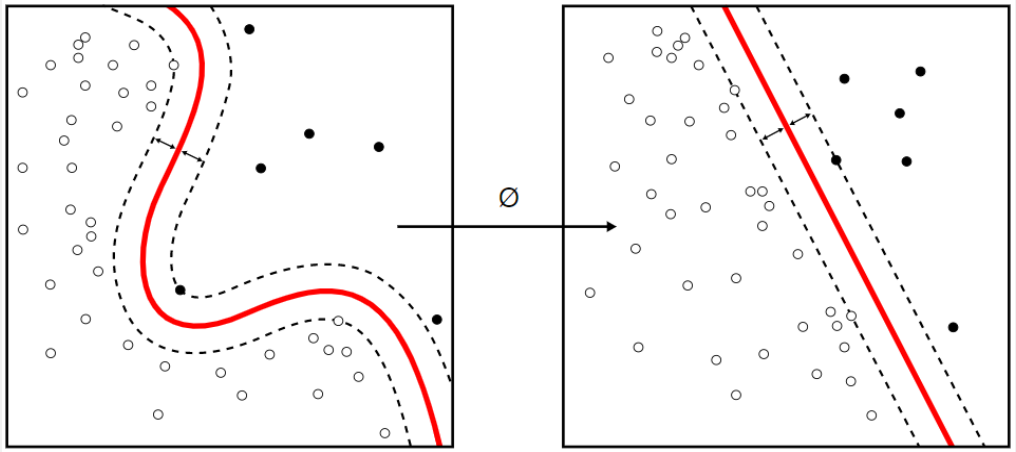
: 오분류에 패널티를 얼마나 줄지 결정하는 척도

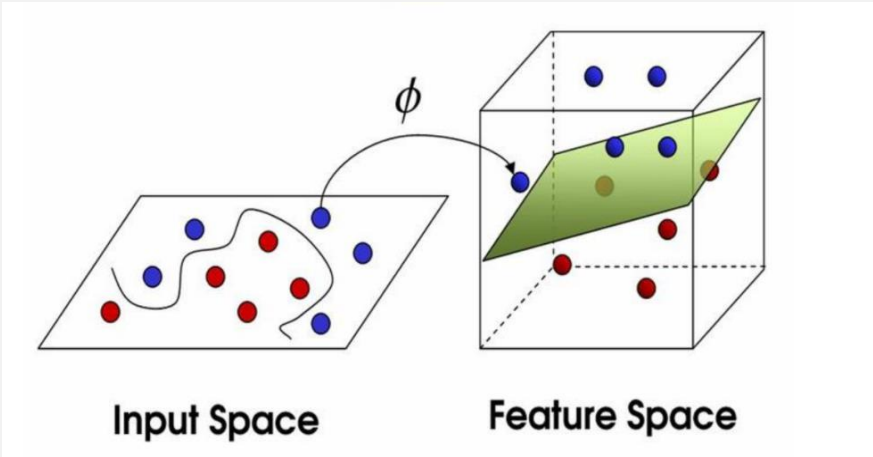


03 Non-Linear SVM

Non-Linear SVM이란

: 선형적으로 분리할 수 없는 데이터셋을 커널 트릭을 통해 초평면을 구하여 데이터를 구별하는 기법





커널 트릭 : 저차원의 데이터를 고차원의 데이터로 매핑하는 작업

Kernel SVM의 종류

① 다항 커널

Kernel = ‘poly’ : 실제로 다항 특성을 만들지 않으면서도 다항식 특성을 많이 추가한 것과 같은 결과를 얻을 수 있는 방법이다.

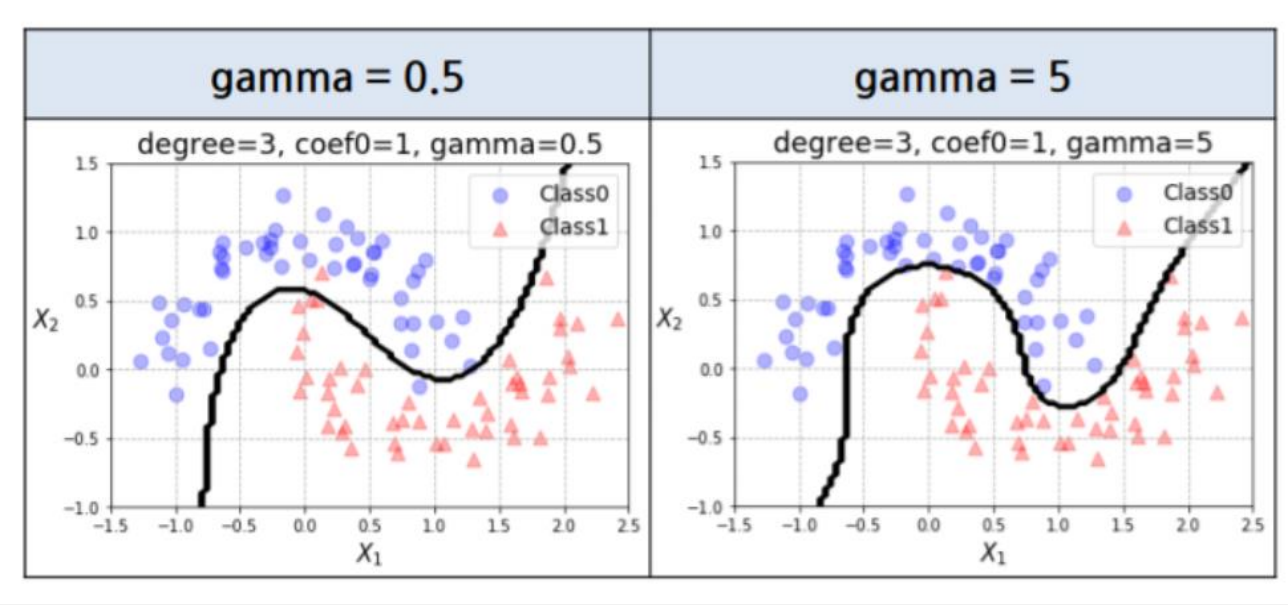
② 가우시안 커널 (= RBF 커널)

Kernel = ‘rbf’ : 진짜로 특성을 늘려서 고차원 공간으로 맵핑(mapping)하는 것은 아니지만,
가우시안 RBF 커널 트릭을 통해 유사도 특성을 많이 추가하는 것과 비슷한 결과를 얻을 수 있다.

03 Non-Linear SVM

하이퍼 파라매터 gamma란

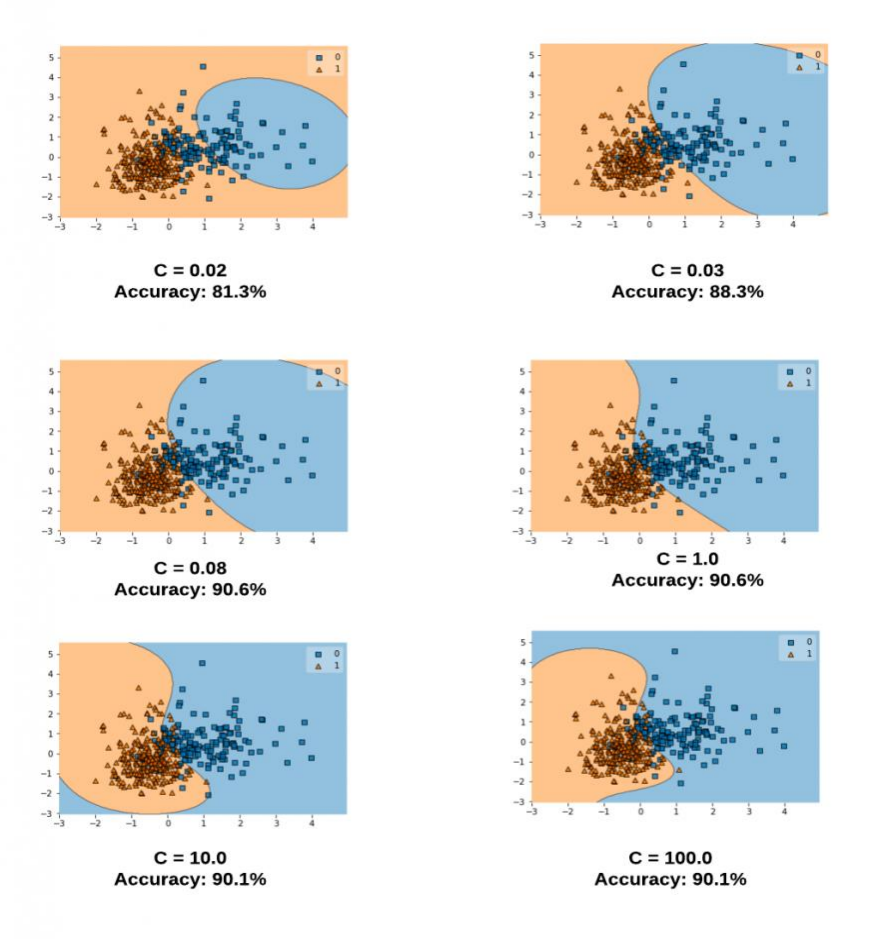
: 얼마나 인접한 값까지 동일한 라벨로 분류하는 지 결정하는 척도 = 결정경계의 곡률을 결정

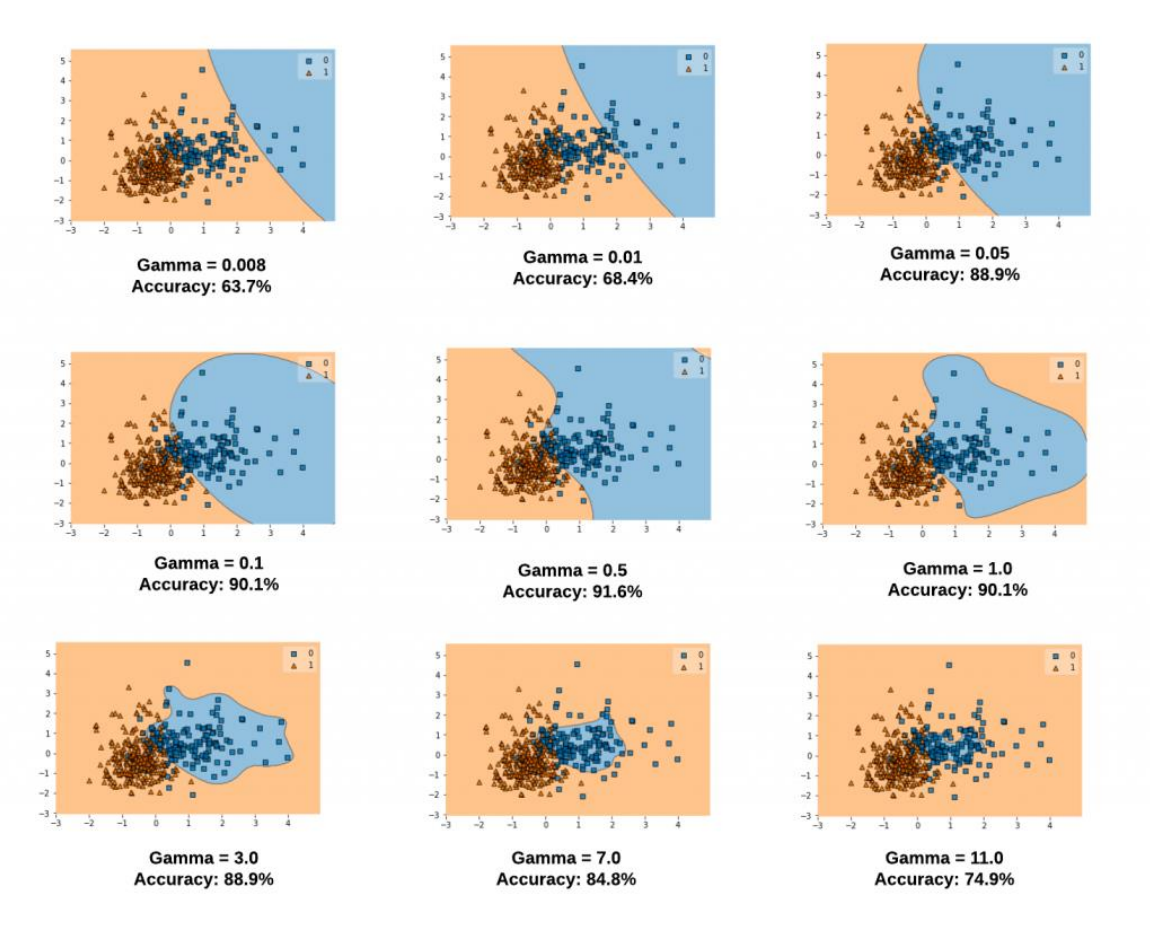


크기	해석
작다	어느정도 인접하면 같은 라벨로 분류
크다	아주 인접한 값만 같은 라벨로 분류

03 Non-Linear SVM

하이퍼 파라매터 c & gamma



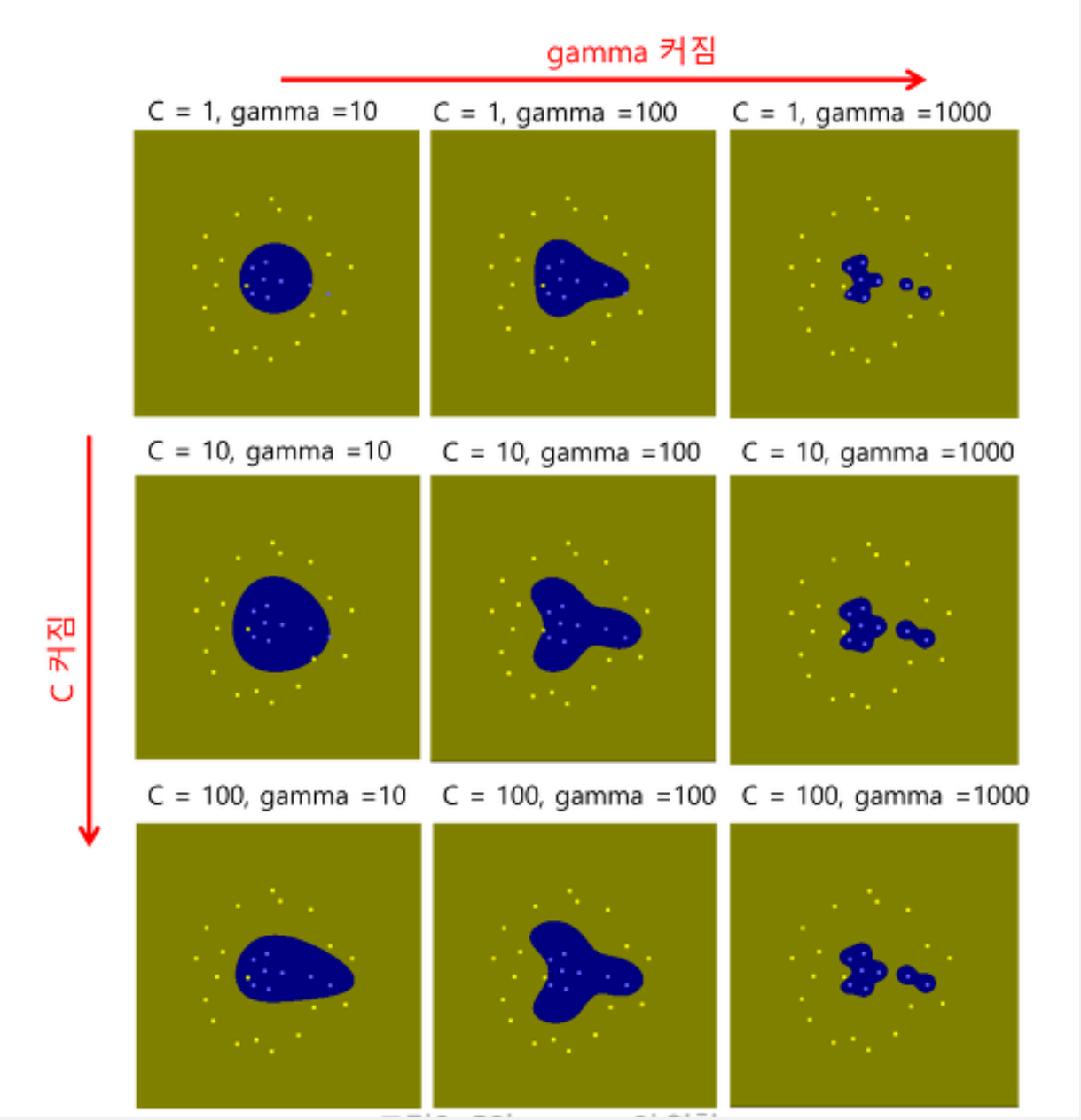


03 Non-Linear SVM

하이퍼 파라매터 c & gamma

- 하이퍼 파라매터 C가 클수록, 오분류를 많이 허용하지 않으므로, 경계가 구불구불한 모습을 보임

- 하이퍼 파라매터 gamma가 클수록 인접한 데이터만 동일 라벨로 인정하므로, 폭이 좁아짐



파이썬

Check-Out

다음주 스터디 발표자를 찾습니다!

감사합니다