BOAZ 분석 15 이재철 , 분석 15 정예림

SVM

Boosting , Stacking

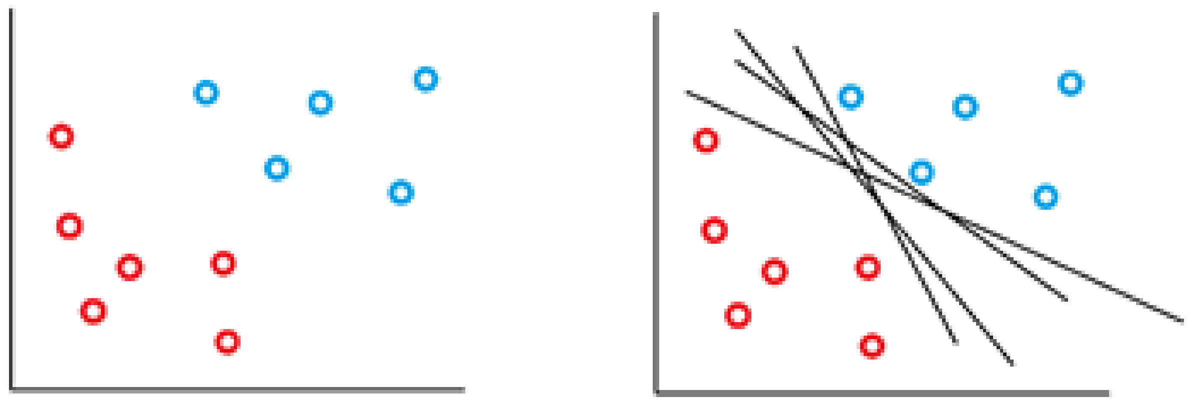
SVM 개요

Support Vector Machine , (*SVM*)

- 지도 학습 모델
- 분류, 회귀 문제
- 탄탄한 이론을 바탕으로 높은 성능을 보여
딥러닝 이전 인기가 많았던 모델

SVM 개요

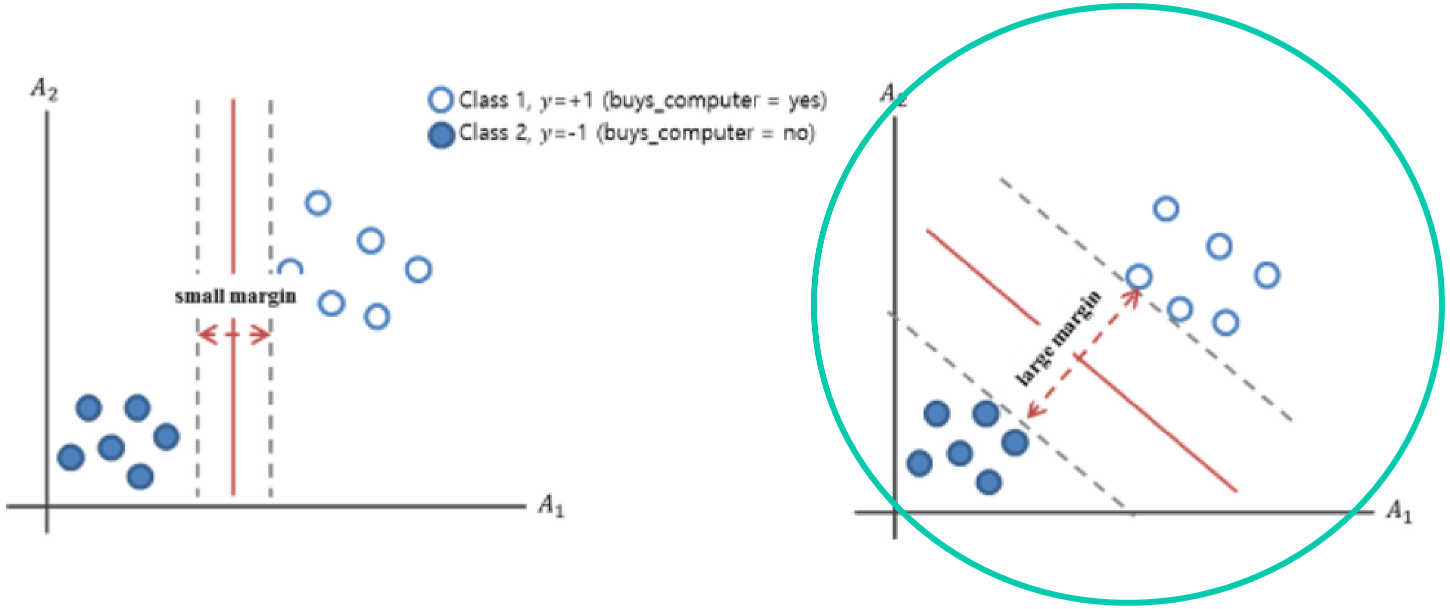
문제 상황



- Two Classification 문제
- 빨간점과 파란점을 나누는 수많은 경계면이 존재함

SVM 개요

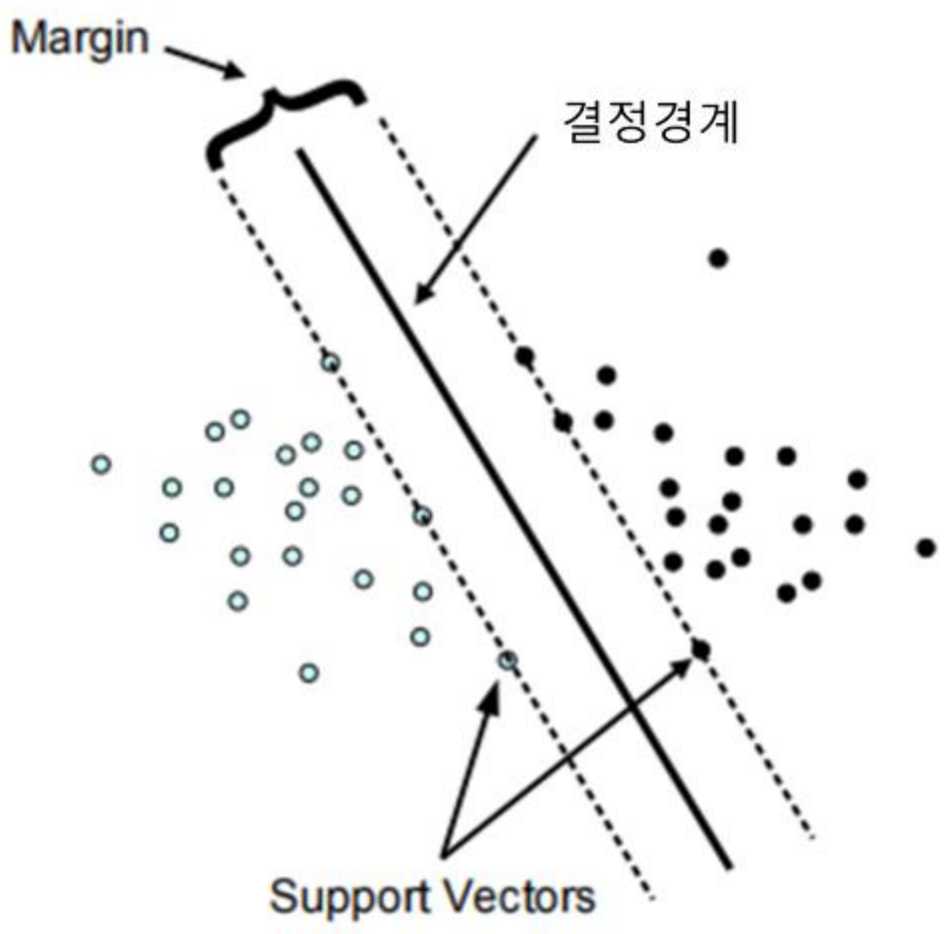
그렇다면, 최적의 경계면은 무엇일까?



"Wide street strategy"

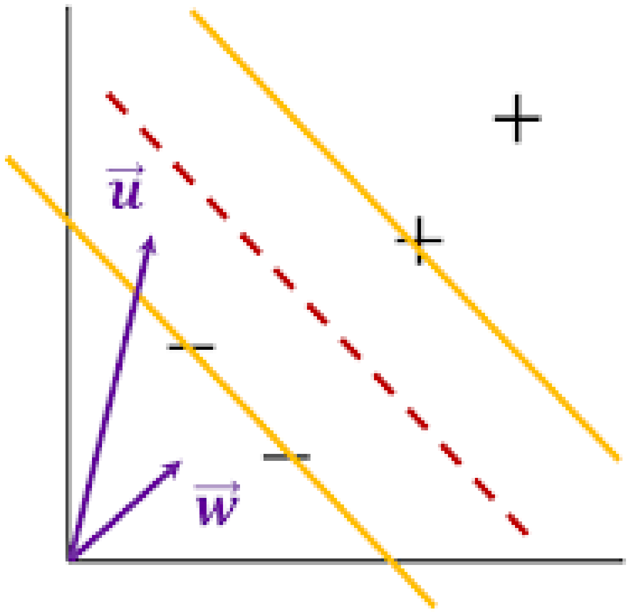
- 마진의 넓이를 최대로 하는 경계면!

SVM 모델



결정 경계 (Decision Boundary)	데이터의 분류 기준이 되는 경계
서포트 벡터 (Support Vector)	학습 데이터 중에서 결정 경계와 가장 가까이에 있는 데이터들의 집합
마진 (Margin)	결정 경계에서 서포트 벡터까지의 거리

SVM 수식



$\vec{w}$: 경계면에 직교하는 벡터

$\vec{u}$: 임의의 벡터

[문제]

$\vec{u}$ 가 경계면을 기준으로 어디에 (+,-) 속하는가?

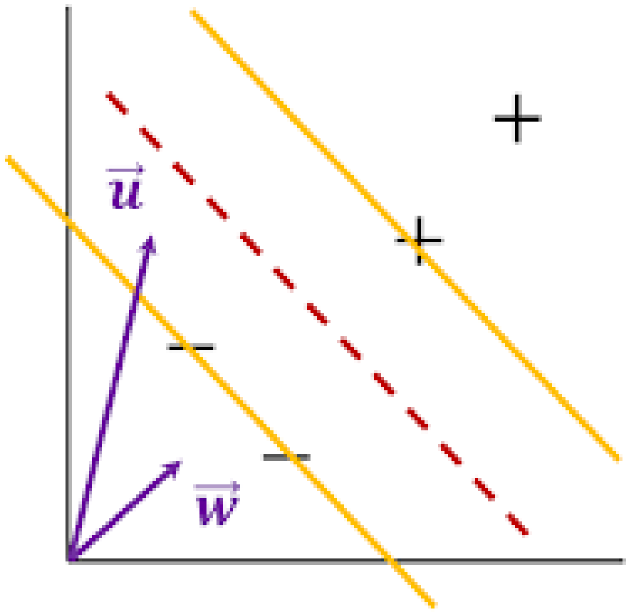
[해결]

$\vec{w}$ 와 $\vec{u}$ 를 내적한 값이 상수 c 보다 큰가?

크다면 , '+' 작다면, '-'

$$\vec{w} \cdot \vec{u} \geq C$$

$$\vec{w} \cdot \vec{u} + b \geq 0 \quad \text{then '+'}$$



$\vec{w}$: 경계면에 직교하는 벡터
x_+: '+' 샘플
x_-: '-' 샘플

1. 어떤 샘플에 대해

$\vec{w} \cdot x_+ + b \geq 1$

$\vec{w} \cdot x_- + b \leq -1$

2. 수학적 편리성을 위해 변수를 추가

$$y_i \begin{cases} 1 & for \quad '+' \\ -1 & for \quad '-' \end{cases}$$

3.1에 곱해서 수식을 하나로 정리

$y_i(\vec{w} \cdot x_i + b) \geq 1$

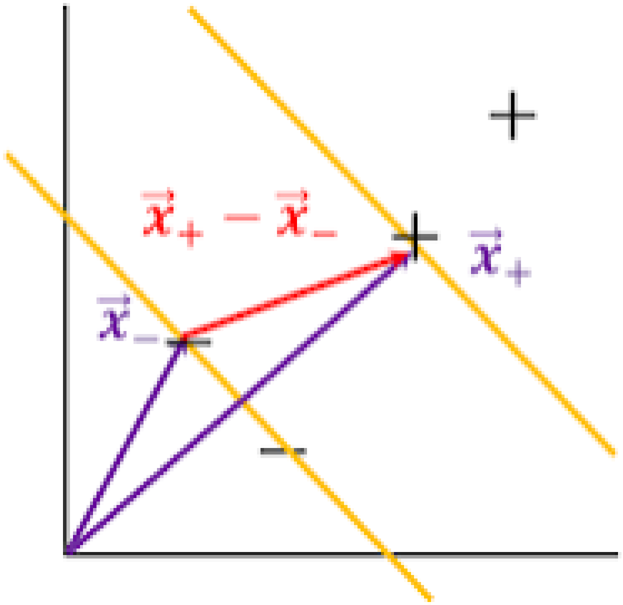
정리하면

$$y_i(\vec{w} \cdot x_i + b) - 1 \geq 0$$

* 등호가 성립할 때 : 샘플이 노란 선에 걸쳐질 때

SVM 수식

우리의 목적 : 마진의 폭을 최대로



마진의 폭 : 두 벡터 $\vec{x}_+$, $\vec{x}_-$ 사이의 폭

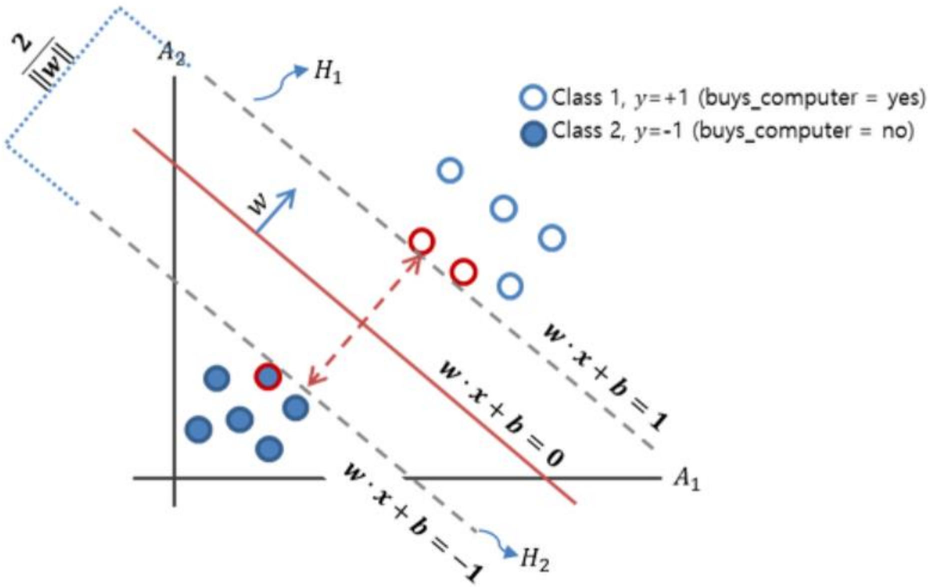
$$WIDTH = (\vec{x}_+ - \vec{x}_-) \cdot \frac{\vec{w}}{||\vec{w}||}$$

이때, $\begin{cases} \vec{w} \cdot x_+ + b \geq 1 \\ \vec{w} \cdot x_- + b \leq -1 \end{cases}$ 를 이용해서 정리

마진은 $\boxed{WIDTH = \frac{2}{||\vec{w}||}}$

$\vec{w}$: 경계면에 직교하는 벡터
$\vec{x}_+$: '+' 샘플
$\vec{x}_-$: '-' 샘플

SVM 수식



마진의 넓이 : $WIDTH = \frac{2}{||\vec{w}||}$

제약식 : $y_i(\vec{w} \cdot x_i + b) - 1 \geq 0$

우리의 목적 : 마진의 폭을 최대로

$$max \frac{1}{||\vec{w}||} \leftrightarrow \min||\vec{w}|| \leftrightarrow min\frac{1}{2}||w||^2$$

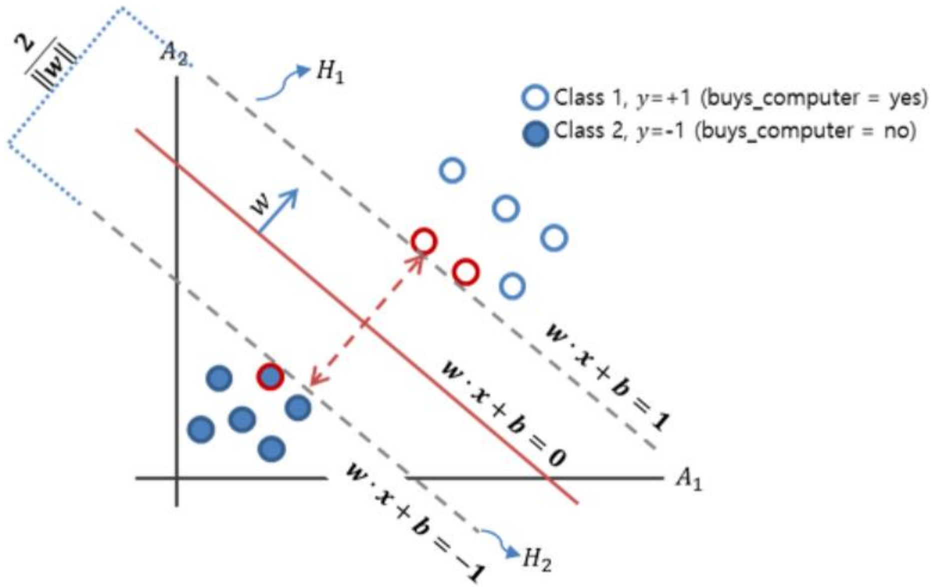
라그랑주 승수법 적용

*제약 조건이 있는 최적화 문제를 제약 조건이 없는 문제로 바꾸기 위해

$$\mathcal{L}(w,b,\alpha) = \frac{1}{2}||\vec{w}||^2 - \sum_{i=1}^{N} \alpha_i[y_i(\vec{w} \cdot \vec{x}_i + b) - 1]$$

$$minimize \quad w.r.t. \quad \vec{w}\ and\ b \qquad maximize \quad w.r.t. \quad \alpha_i \geq 0 \quad \forall i$$

SVM 수식



마진의 넓이 : $WIDTH = \frac{2}{||\vec{w}||}$

제약식 : $y_i(\vec{w} \cdot x_i + b) - 1 \geq 0$

라그랑주 승수법 적용

*제약 조건이 있는 최적화 문제를 제약 조건이 없는 문제로 바꾸기 위해

$$\mathcal{L}(w, b, \alpha) = \frac{1}{2}||\vec{w}||^2 - \sum_{i=1}^{N} \alpha_i [y_i(\vec{w} \cdot \vec{x}_i + b) - 1]$$

$minimize \quad w.r.t. \quad \vec{w}\ and\ b \qquad maximize \quad w.r.t. \quad \alpha_i \geq 0 \quad \forall i$

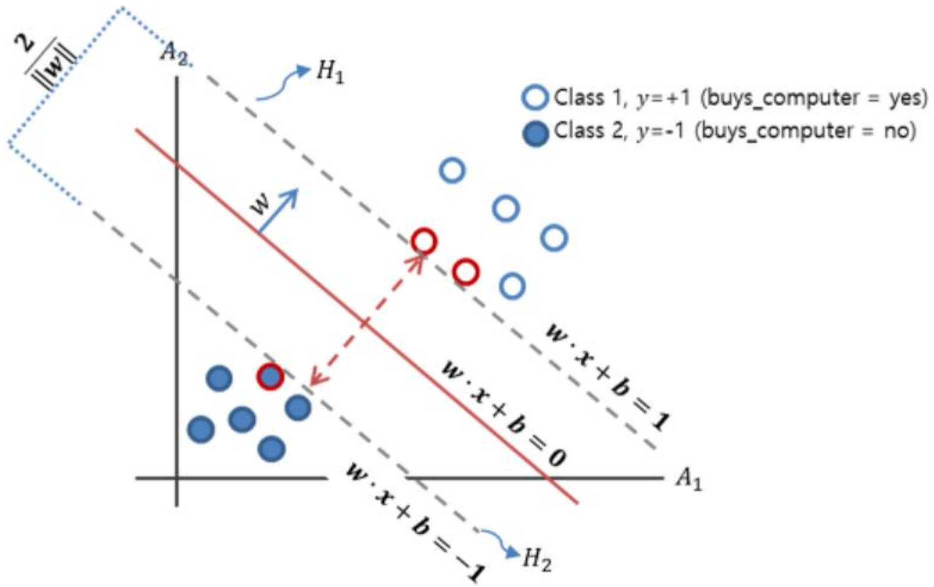
$\vec{w}$ 와 b 에 대해 각각 편미분

$$\nabla_{\vec{w}} \mathcal{L} = \vec{w} - \sum_i \alpha_i y_i \vec{x}_i = 0$$

$$\nabla_b \mathcal{L} = -\sum_i \alpha_i y_i = 0$$

$$\vec{w} = \sum_i \alpha_i y_i \vec{x}_i$$

SVM 수식



마진의 넓이 : $WIDTH = \frac{2}{||\vec{w}||}$

제약식 : $y_i(\vec{w} \cdot x_i + b) - 1 \geq 0$

α 에 대한 maximization 문제로 정리

$$\sum_{i=1}^{N} \alpha_i + \frac{1}{2}||\vec{w}||^2 - \sum_{i=1}^{N} \alpha_i y_i \vec{w}^T \vec{x}_i \iff \sum_{i=1}^{N} \alpha_i + \frac{1}{2}\vec{w}^T\vec{w} - \vec{w}^T\vec{w}$$
$$\iff \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\vec{w}^T\vec{w}$$
$$\iff \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\alpha_j y_i y_j \vec{x}_i^T \vec{x}_j$$
$$\iff \mathcal{L}(\alpha).$$

1. α를 구하면 $\vec{w}$ 를 구할 수 있음

2. α값이 0이 아니라는 것은 해당 $\vec{x}$가 경계선을 정하는 샘플

즉, Support Vector 라는 뜻

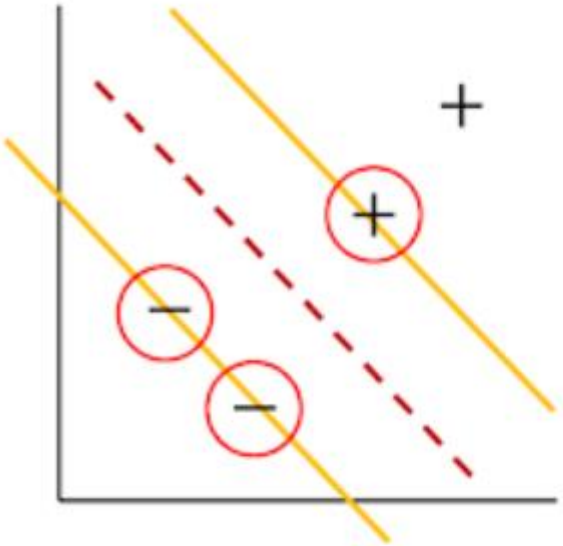
3. $L(\alpha)$ 식 의 성질에 의해

" SVM 으로 구한 해는 항상 (현재 SVM)이 줄 수 있는 최적의 해"

임을 이론적으로 보장. -> local minima에 빠지지 않음

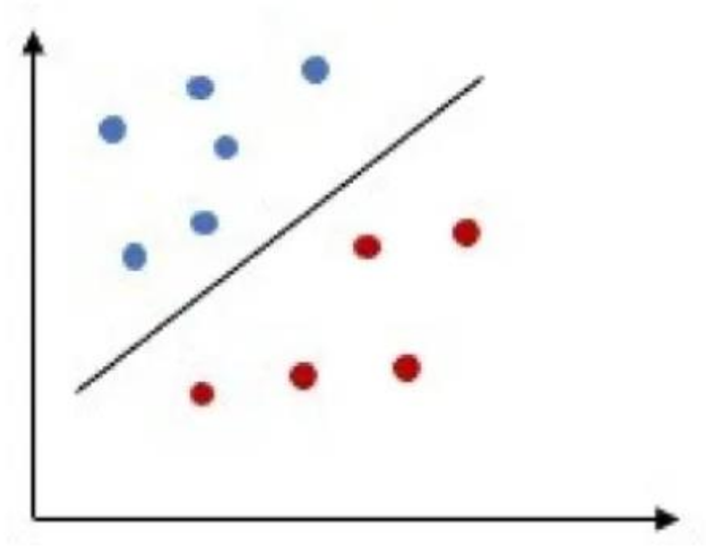
SVM 수식

결론



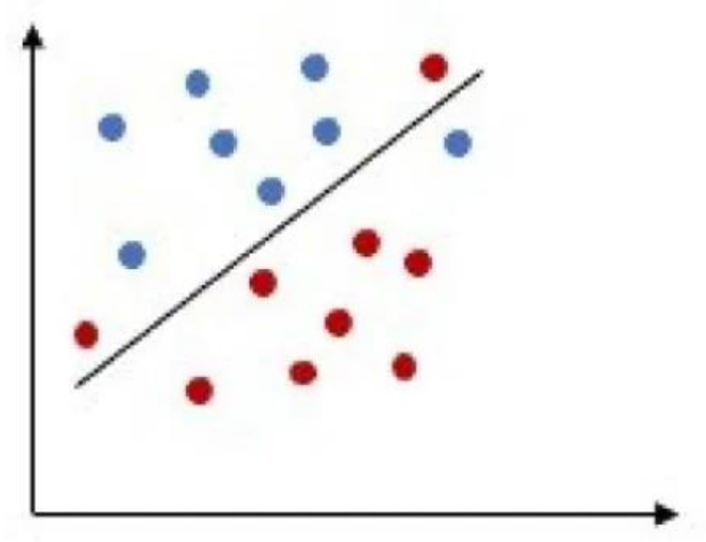
support vector들로 정해진 decision boundary가 가장 최적의 boundary이며,
새로운 샘플이 들어왔을 때 일반화를 가장 잘 할 수 있는
decision rule을 찾을 수 있게 된 것

SVM 종류



하드 마진 SVM (Hard Margin SVM)

- 오분류를 허용하지 않는 SVM
- 노이즈로 인해 최적의 결정 경계를 잘 못 구하거나, 못 찾을 수 있음



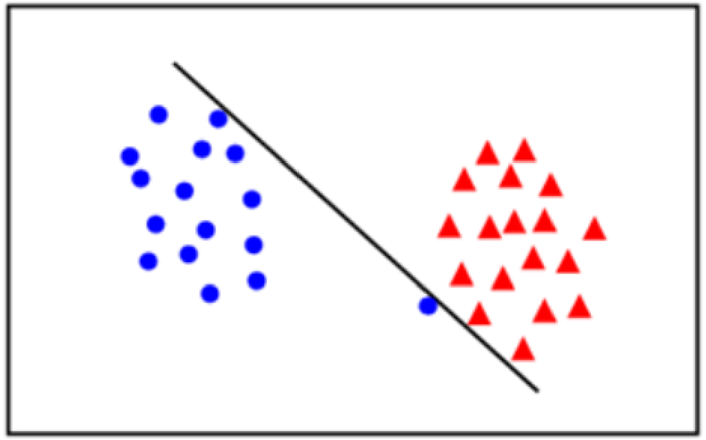
소프트 마진 SVM (Soft Margin SVM)

- 오분류를 허용하는 SVM
- 하드 마진 SVM은 적용하기가 어려우므로 어느 정도의 오류를 허용하는 소프트 마진 SVM을 주로 이용

Soft Margin SVM – 슬랙 변수

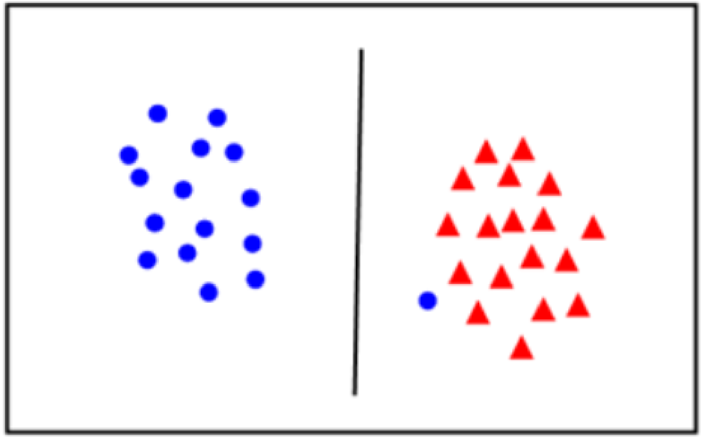
둘 중 최적의 경우는?

Hard



마진 ↓ , 오분류 X

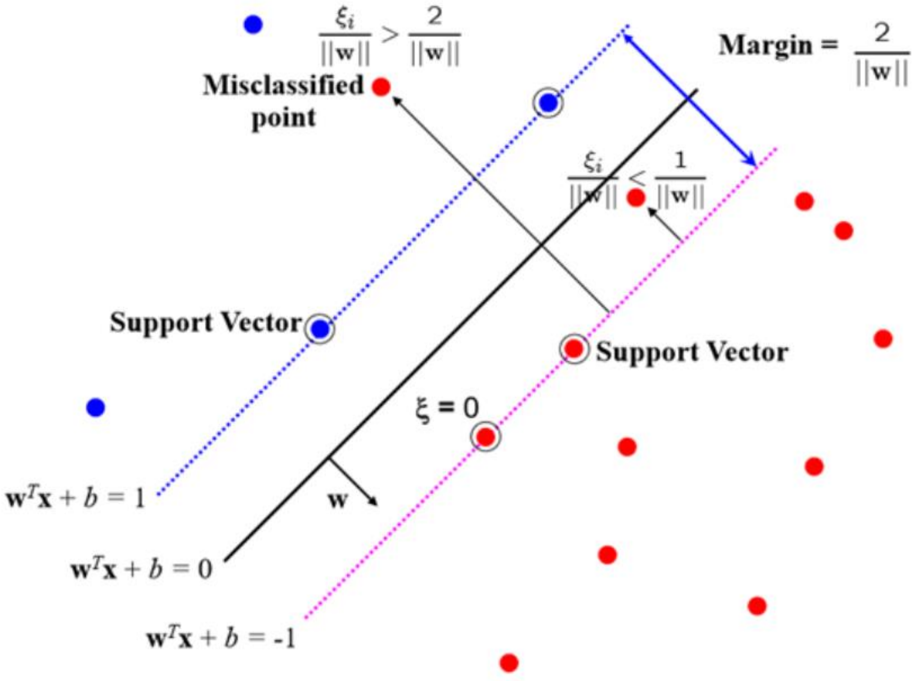
Soft



마진 ↑ , 오분류 O

일반적으로 마진과 학습 오류의 개수는 Trade-Off 관계임

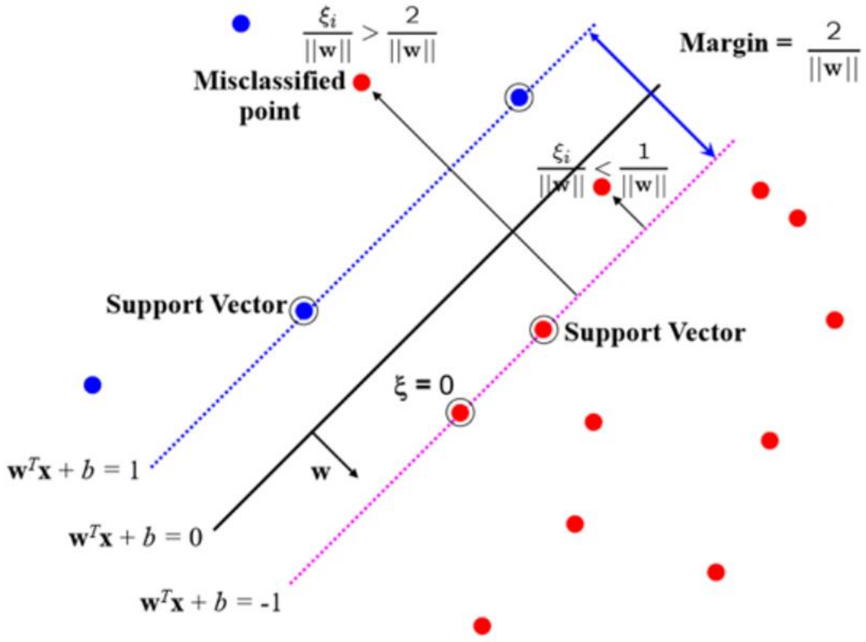
Soft Margin SVM – 슬랙 변수



슬랙 변수 (Slack Variables)

데이터의 완벽한 분리가 불가능 할 때, 선형적 분류를 위해 허용된 오차를 위한 변수

C-SVM



C – SVM

*중치

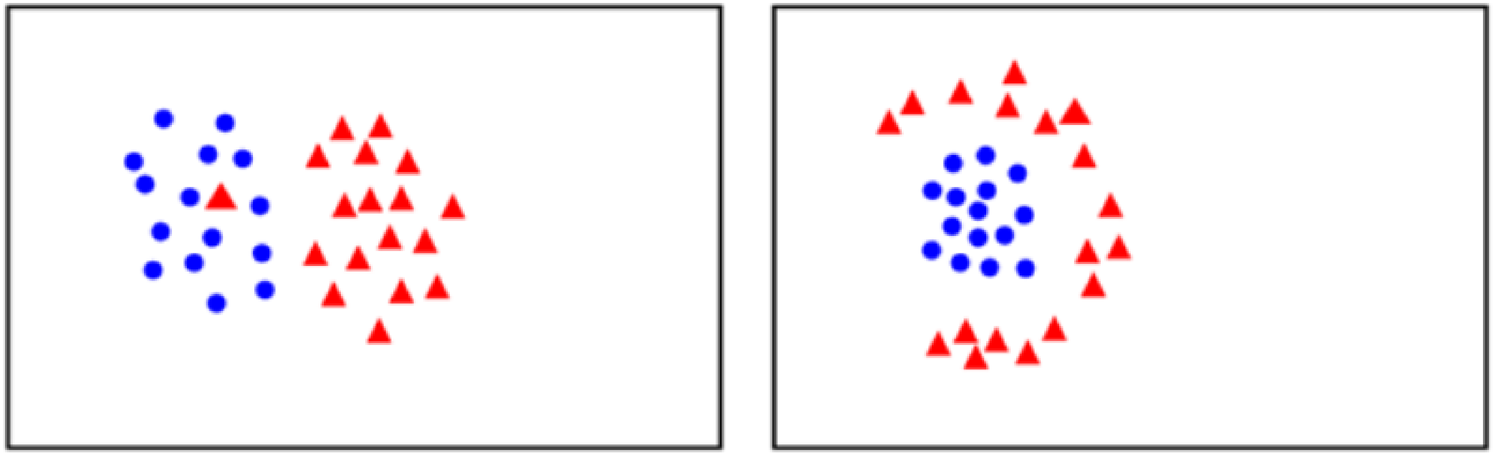
- Soft Margin SVM
- 엡실론만큼 패널티를 부과
- C의 크기로 마진의 폭을 조절

목적식 변화

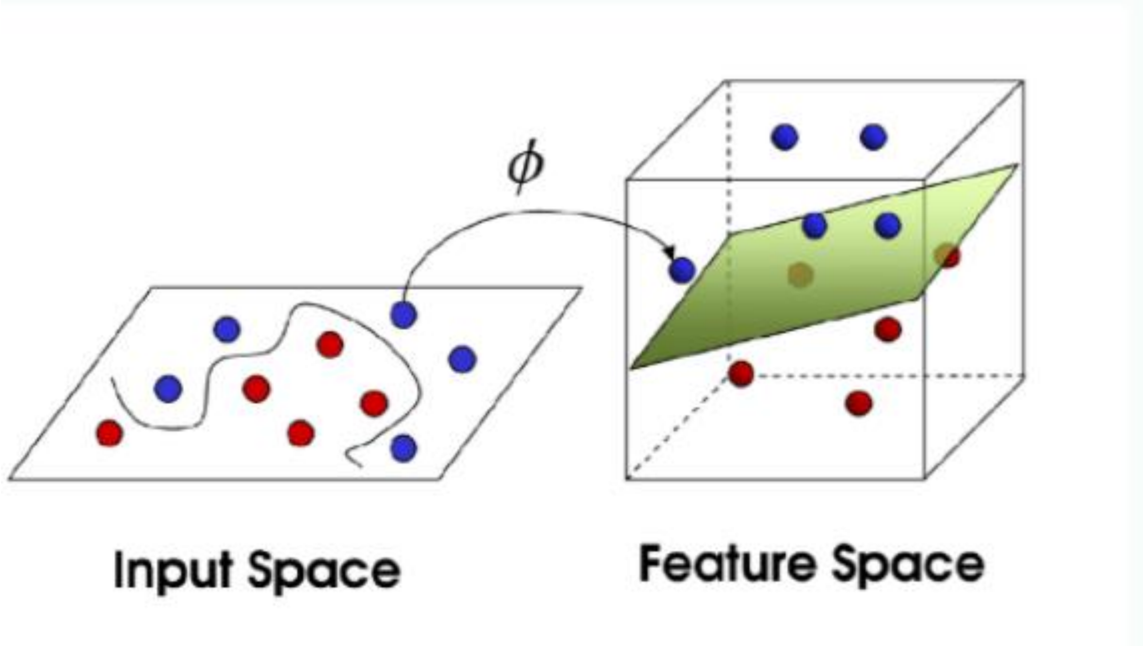
$$min\frac{1}{2}||w||_2^2 + C\sum_{i=1}^{n}\varepsilon_i$$

데이터 분류가 비선형일 경우

not
linearly
separable



Kernel Trick (비선형일 경우)



" Kernel trick "

비선형인 경우 저차원 공간을 고차원 공간으로 매핑하는 방법
"커널 함수"를 이용하여 고차원 공간으로 매핑하는 경우에
증가하는 연산량의 문제를 해결하는 기법

커널 함수

주로 사용하는 커널 함수

Type of Kernel	Inner product kernel $K(\vec{x},\vec{x}_i),\ i=1,2,\ldots,N$	Comments
Polynomial Kernel	$K(\vec{x},\vec{x}_i)=(\vec{x}^T\vec{x}_i+\theta)^d$	Power p and threshold θ is specified a priori by the user
Gaussian Kernel	$K(\vec{x},\vec{x}_i)=e^{-\frac{1}{2\sigma^2}\Vert\vec{x}-\vec{x}_i\Vert^2}$	Width σ^2 is specified a priori by the user
Sigmoid Kernel	$K(\vec{x},\vec{x}_i)=tanh(\eta\vec{x}\vec{x}_i+\theta)$	Mercer's Theorem is satisfied only for some values of η and θ
Kernels for Sets	$K(\chi,\chi')=\sum_{i=1}^{N_\chi}\sum_{j=1}^{N_{\chi'}}k(x_i,x'_j)$	Where $k(x_i,x'_j)$ is a kernel on elements in the sets χ, χ'
Spectrum Kernel for strings	count number of substrings in common	It is a kernel, since it is a dot product between vectors of indicators of all the substrings.

이중, 가우시안 커널 을 가장 많이 사용함

가우시안 커널 (= RBF , Radial Basis Fuction)

SVM 사용하기

Sklearn 라이브러리는 다양한 SVM 모델을 지원함

상황	모델	
분류 문제	SVC	Classification의 C
	LinearSVC	SVC의 linear kernel 특화 버전
회귀 문제	SVR	Regression의 R
	LinearSVR	SVR의 linear kernel 특화 버전

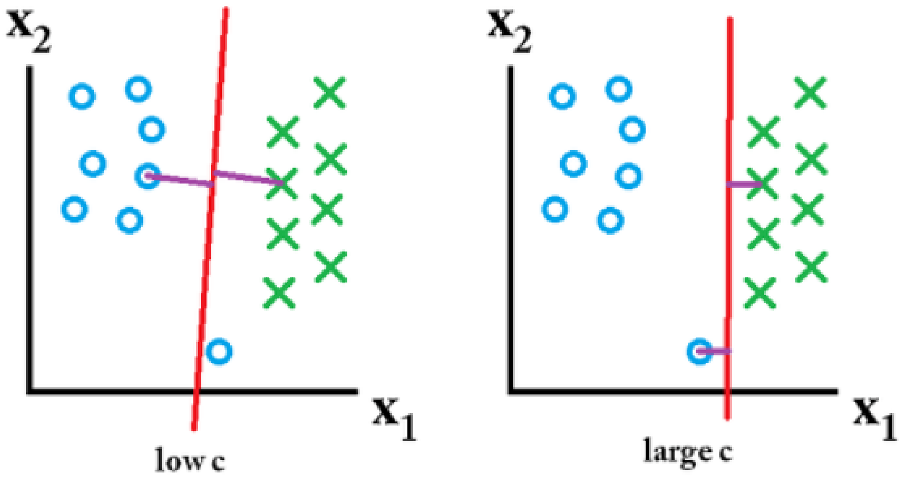
SVM 사용하기

SVM의 매개변수

C , gamma

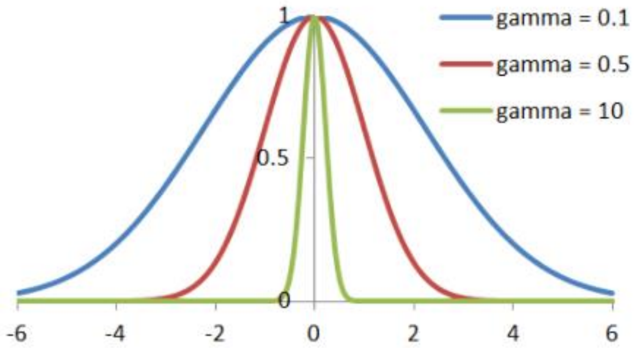
SVM의 매개변수 - C

C : 소프트 마진 SVM의 슬랙변수



C가 너무 낮으면 과소 적합
C가 너무 높으면 과대 적합

gamma : 가우시안 함수의 표준편차와 관련 있는 매개변수



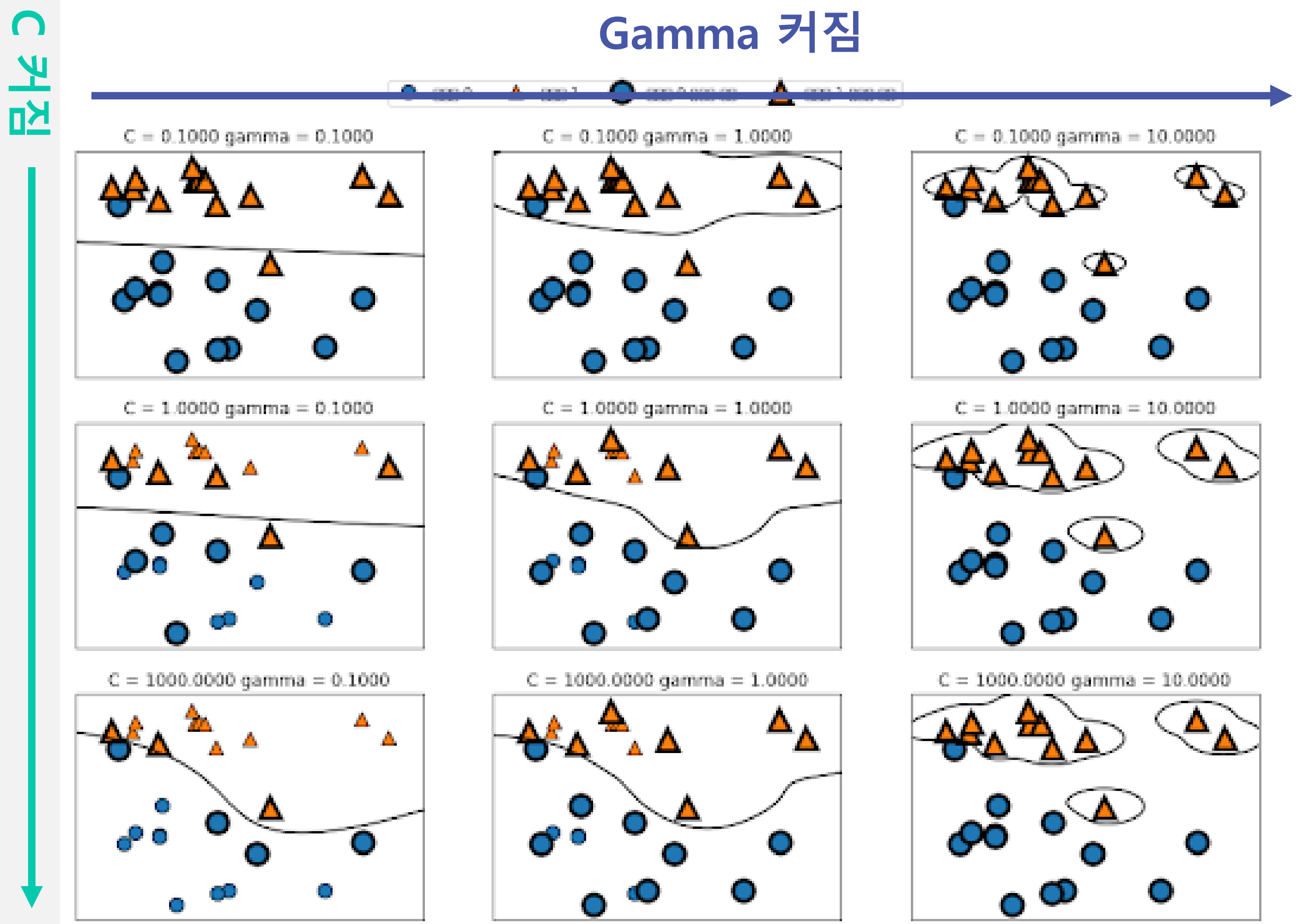
Gamma 는 RBF 커널에서
하나의 데이터 샘플이 영향력을 행사하는 거리를 나타냄

Gamma가 크면 데이터 포인터들이 영향력을 행사하는 거리가 짧아지고
Gamma가 낮으면 영향력을 행사하는 거리가 커짐

Gamma가 너무 낮으면 과소 적합
Gamma가 너무 크면 과대 적합

SVM 사용하기

SVM의 매개변수 조절



Gamma가 동일한 경우
C가 커질 수록 오분류 인정 X

C 가 동일한 경우
gamma가 커질 수록 곡률 높음

SVM 사용하기

SVM의 매개변수 조절

C, gamma

C 커짐

Gamma 커짐

$2^{-5}, 2^{-15}$	$2^{-3}, 2^{-15}$		$2^{11}, 2^{-15}$	$2^{13}, 2^{-15}$
$2^{-5}, 2^{-13}$	$2^{-3}, 2^{-13}$		$2^{11}, 2^{-13}$	$2^{13}, 2^{-13}$
$2^{-5}, 2^{-11}$	$2^{-3}, 2^{-11}$		$2^{11}, 2^{-11}$	$2^{13}, 2^{-11}$
$2^{-5}, 2^{-9}$	$2^{-3}, 2^{-9}$		$2^{11}, 2^{-9}$	$2^{13}, 2^{-9}$
$2^{-5}, 2^{-7}$	$2^{-3}, 2^{-7}$	...	$2^{11}, 2^{-7}$	$2^{13}, 2^{-7}$
$2^{-5}, 2^{-5}$	$2^{-3}, 2^{-5}$		$2^{11}, 2^{-5}$	$2^{13}, 2^{-5}$
$2^{-5}, 2^{-3}$	$2^{-3}, 2^{-3}$		$2^{11}, 2^{-3}$	$2^{13}, 2^{-3}$
$2^{-5}, 2^{-1}$	$2^{-3}, 2^{-1}$		$2^{11}, 2^{-1}$	$2^{13}, 2^{-1}$
$2^{-5}, 2^{1}$	$2^{-3}, 2^{1}$		$2^{11}, 2^{1}$	$2^{13}, 2^{1}$
$2^{-5}, 2^{3}$	$2^{-3}, 2^{3}$		$2^{11}, 2^{3}$	$2^{13}, 2^{3}$

Grid search

매개변수들의 여러 조합들을 테스트 해서 가장 좋은 성능을 내는 매개변수를 찾아내는 것

SVM 참고 자료

SVM 수식 이해 참고자료	**Youtube - MIT OpenCourseWare** ★★★★★ https://www.youtube.com/watch?v=_PwhiWxHK8o&t=6s **Blog - JaeJun Yoon's Playground** ★★★★☆ http://jaejunyoo.blogspot.com/2018/01/support-vector-machine-1.html
SVM 종류	**Blog – 라온피플** ★★☆☆☆ https://blog.naver.com/laonple/220847975603 **Blog – 귀퉁이 서재** ★★★☆☆ https://bkshin.tistory.com/entry/%EB%A8%B8%EC%8B%A0%EB%9F%AC%EB%8B%9D-2%EC%84%9C%ED%8F%AC%ED%8A%B8-%EB%B2%A1%ED%84%B0-%EB%A8%B8%EC%8B%A0-SVM
SVM 사용하기	**Blog – bskyVision** ★★★☆☆ https://blueskyvision.tistory.com/163 **Blog – 코딩생활 커딩왕** ★★★★★ https://chancoding.tistory.com/67