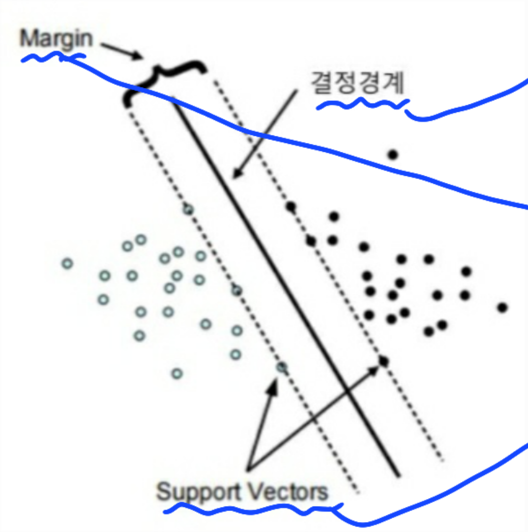
SVM(Support Vector Machine)

: Large Margin Classification이라고도 함

: SVM 분류기는 클래스 사이에 가장 폭이 넓은 마진을 찾는 것

: 특성의 스케일에 민감



데이터의 분류 기준이 되는 경계

결정 경계에서 서포트 벡터까지의 거리

학습 데이터중 결정 경계와 가장 가까이 있는 데이터들의 집합

하드 마진 SVM (Hard Margin SVM)

- 오분류를 허용하지 않는 SVM → 마진오류 엄격
- 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

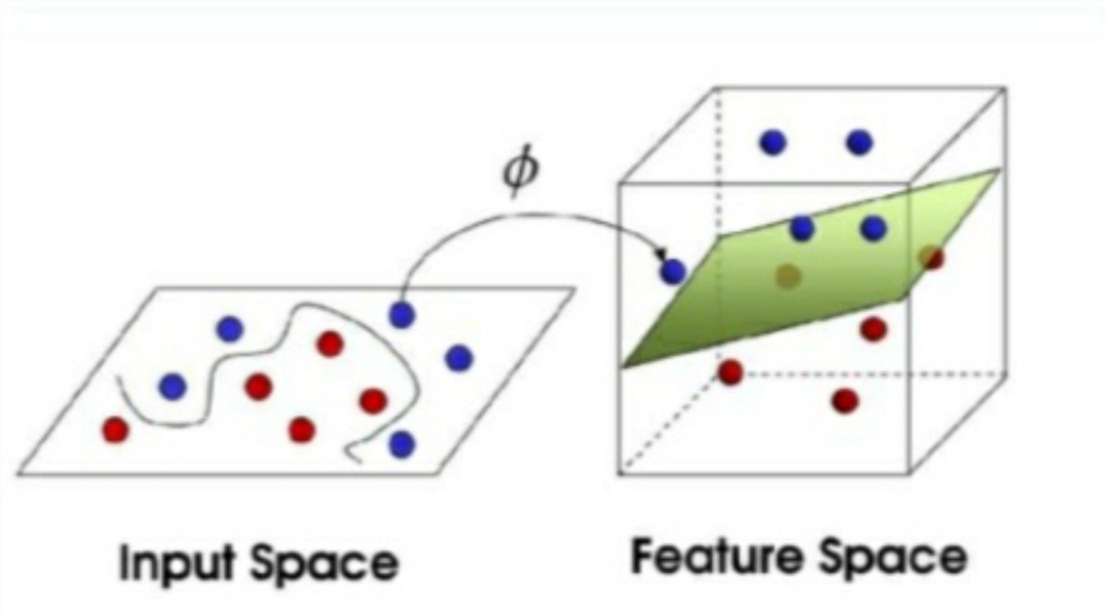
마진오류 결정파라미터 : C → C ↑ → 마진오류 ⇓ ⇒ Hard margin classification
C ↓ → 마진오류 ⇑ ⇒ Soft Margin classification

If svm overfit ⇒ C ⇓

비선형 SVM

Kernel Trick을 이용하여 매핑

-> 고차원 공간으로 매핑하는 경우에 증가하는 연산량의 문제를 해결



SVM kernel ⇒ Poly 로 설정

coef0 ⇒ 다항식 커널에 있는 상수항 — coef0 ⇑
⇒ 고차항의 영향 ⇓

유사도 특성

: 각 샘플이 특정 랜드마크와 얼마나 닮았는지 측정하는 것

- 가우시안 RBF (유사도 함수)

$\phi_\gamma(x, \ell) = \exp(-\gamma \|x-\ell\|^2)$ ⇒ 0~1 사이 종 모양 (0은 랜드마크와 아주 멀리 떨어진 경우)
(1은 랜드마크와 같은 위치인 경우)

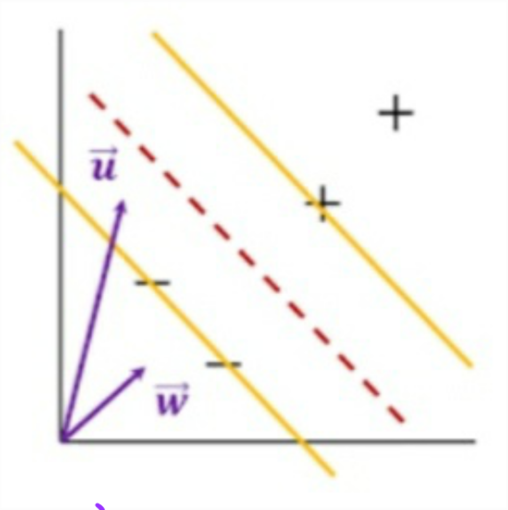
가우시안 RBF kernel

kernel = rbf , gamma
- γ ⇑ → 종모양(RBF) 좁아짐 ⇒ 샘플의 영향 범위 작아짐
- γ ⇓ → 종모양(RBF) 넓어짐 ⇒ 결정경계 부드러워짐

⇓

규제역할!! ⇒ SVM Overfit ⇒ γ ⇓
〃 underfit ⇒ γ ⇑

SVM 수식



$\vec{u}$ 가 경계면을 기준으로 어디 (+, −)에 속하는가?

⇒ $\vec{w}\cdot\vec{u} \ge C$ (상수) ⇒ + else −

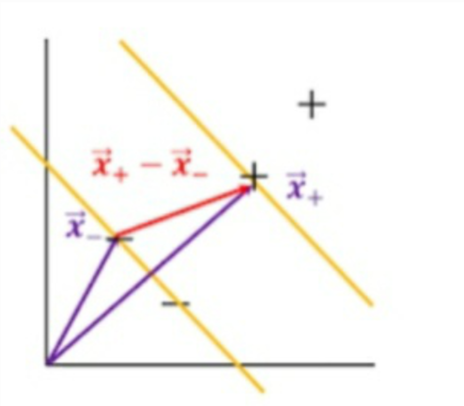
$$\boxed{\begin{aligned} \vec{w}\cdot x_+ + b &\ge 1 \\ \vec{w}\cdot x_- + b &\le -1 \end{aligned}}$$

(x_+: + 샘플, x_-: − 샘플)

⇒ $\boxed{y_i \begin{cases} 1 & \text{for } + \\ -1 & \text{for } - \end{cases}}$ ⇒ $\underbrace{y_i(\vec{w}\cdot x_i + b) - 1 \ge 0}_{\text{s.t (제약식)}}$

$\vec{w}$: 경계면에 직교하는 벡터

$\vec{u}$: 임의의 벡터



목적: Margin 폭을 최대로

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

$$\begin{aligned} &\vec{w}\cdot x_+ + b \ge 1 \\ -\ &\vec{w}\cdot x_- + b \le -1 \\ \hline &\vec{w}(x_+ - x_-) \ge 2 \end{aligned}$$

→ WIDTH = $\frac{2}{\|\vec{w}\|}$

라그랑지 승수법 이용

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

$\vec{w}$에 대해 편미분 : $\nabla_{\vec{w}} L = \vec{w} - \sum \alpha_i y_i \vec{x}_i = 0$

b에 대해 편미분 : $\nabla_b L = -\sum \alpha_i y_i = 0$

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

α에 대한 maximization 문제로 정리

1. α를 구하면 $\vec{w}$ 구할 수 있음
2. α ≠ 0 , $\vec{x}$ 가 경계선을 정하는 샘플 즉, Support Vector
3. SVM으로 구한 해는 라그랑지 최적화 이론에 의해 최적의 해 증명 (Local Minima에 안빠짐)