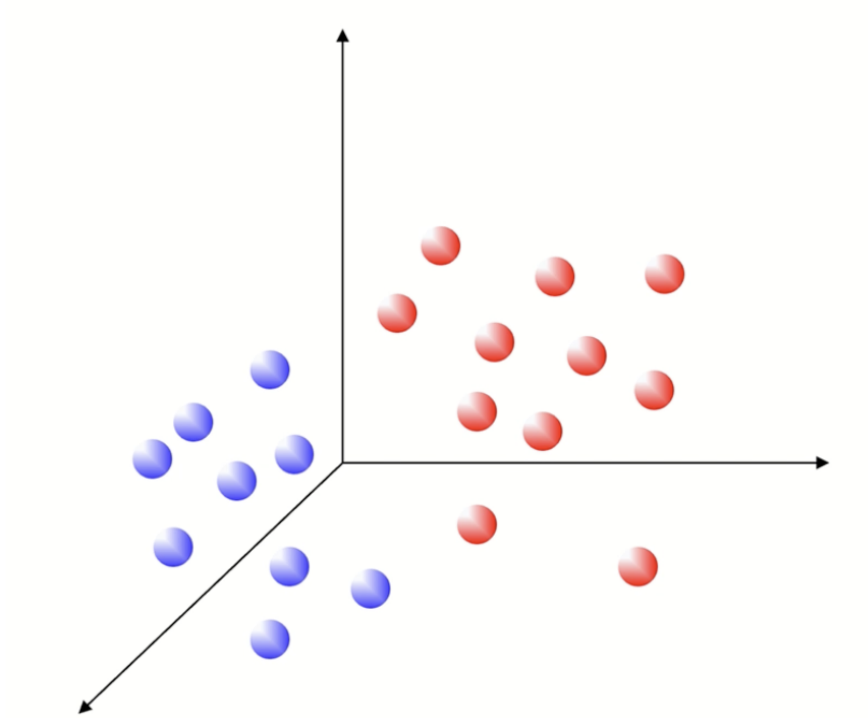
SVM

SVM 전반적인 설명

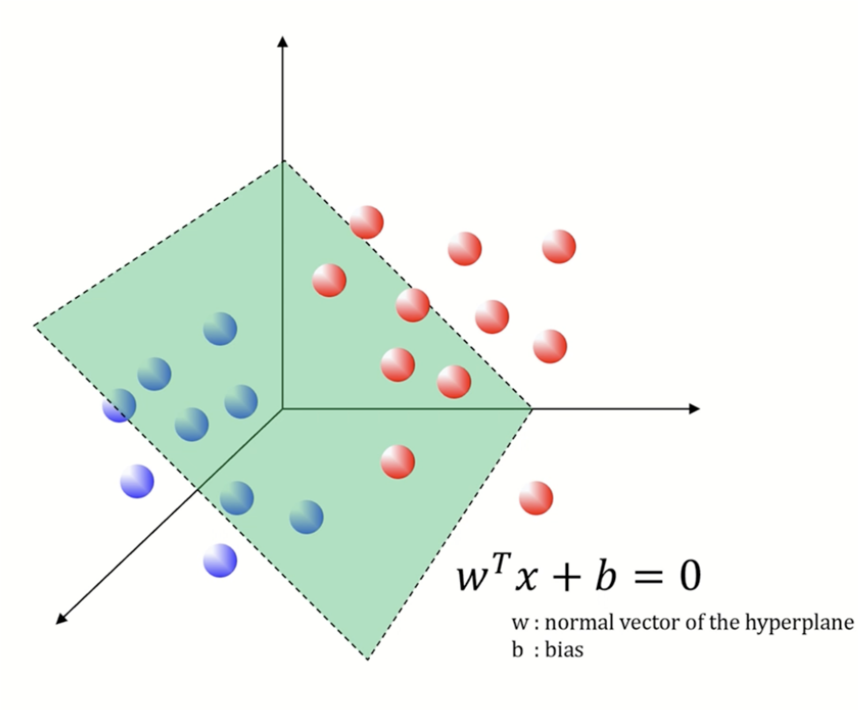
svm : support Vector Machine

- 고차원 데이터의 분류문제에 좋은 성능
- 트레이드오프(generalization ability, training data) 관계에서 generalization ability 증가시키는 방향
- Statistical learning theory에 근거

SVM 상황



→ 이진분류문제



→ hyperplane을 찾자!

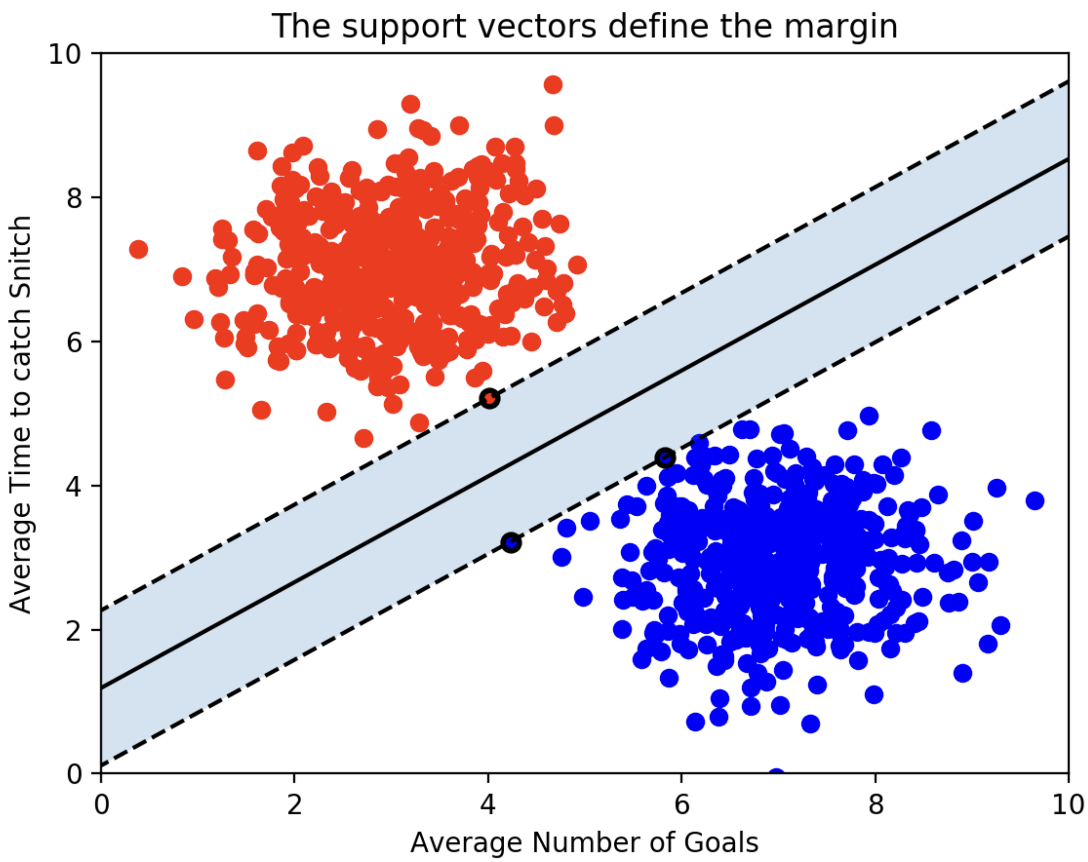
그런데 너무 많다...

svm 기준 | Margin

바로 Margin!

마진을 최대화할 수 있는 hyperplane을 찾자

→ generalization error를 최소화



Margin :

각 클래스에서 가장 가까운 관측치 사이의 거리

W로 표현 가능 (Wx + b)

1) Plus-plane

빨간색 클래스가 y=1

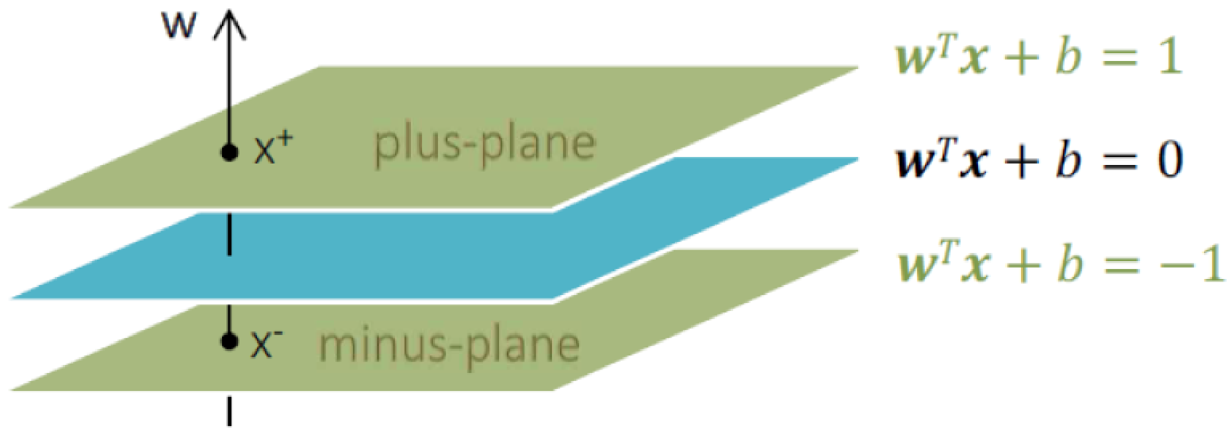
$$W^T X_i + b \geq 1$$

2) Minus-plane

파란색 클래스가 y=-1

$$W^T X_i + b \leq -1$$

이 2개 사이의 거리가 Margin이고 이 마진을 최대화하는 hyperplane을 찾자



최적의 decision boundary

$$W^T X_i + b = 0$$

그리고 이 둘의 관계는

$$x^+ = x^- + \lambda w$$

즉, 평행이동관계

수학적으로 보는 Margin

λ 구하기

$w^T x^+ + b = 1$

$\Rightarrow w^T(x^- + \lambda w) + b = 1$

$\Rightarrow w^T x^- + b + \lambda w^T w = 1$

여기서 $w^T x^- + b$ = -1

$\Rightarrow -1 + \lambda w^T w = 1$

$\lambda = \frac{2}{w^T w}$

⚠️잊지말자 우리는 W에 주목!⚠️

Margin 구하기

$$
\begin{aligned}
&Margin \\
&\quad = distance(x^+, x^-) \\
&\quad = ||x^+ - x^-||_2 \\
&\quad = ||x^- + \lambda w - x^-||_2 \\
&\quad = ||\lambda w||_2 \\
&\quad = \lambda\sqrt{w^T w} \\
&\quad = \frac{2}{\sqrt{w^T w}} \\
&\quad = \frac{2}{||w||_2}
\end{aligned}
$$

결론!

max margin = max $\frac{2}{||w||_2}$ ⇒ min $\frac{||w||_2}{2}$

목적식 : $minimize \frac{||w||_2^2}{2}$

제약식 : $y_i(w^T x_i + b) \geq 1, i = 1, 2, 3, ..., n$

Lagrangian multiplier를 이용하여 Lagrangian Primal

$$max_a min_{w,b} L(w, b, \alpha) = \frac{1}{2}||w||_2^2 - \sum_{i=1}^{n} \alpha_i (y_i(w^T x_i + b) - 1)$$

우선 **min**문제 해결 → 미분 = 0

- W에 대한 미분 ⇒ $w = \sum_{i=1}^{n} \alpha_i y_i x_i$
- b에 대한 미분 ⇒ $\sum_{i=1}^{n} \alpha_i y_i = 0$

$$\frac{1}{2}||w||_2^2 = \frac{1}{2}\sum_{i=1}^{n} \alpha_i y_i (\sum_{i=1}^{n} \alpha_i y_i x_i^T x_i) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$\sum_{i=1}^{n} \alpha_i (y_i(w^T x_i + b) - 1) = \sum_{i=1}^{n} \alpha_i y_i (w^T x_i + b) + \sum_{i=1}^{n} \alpha_i = -\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_{i=1}^{n} \alpha_i$$

*** 이제 α만 구하면 W, b 알 수 있다**

이제 **max**문제를 해결한다. | Lagrangian dual solution

$\sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$

이 문제를 풀기 위해서는 KKT conditions을 만족해야 한다.

① Stationarity

$$\frac{\partial \mathcal{L}(w,b,\alpha)}{\partial w} = 0 \implies w = \sum_{i=1}^{n} \alpha_i y_i x_i \qquad \frac{\partial \mathcal{L}(w,b,\alpha)}{\partial b} = 0 \implies \sum_{i=1}^{n} \alpha_i y_i = 0$$

② Primal feasibility $\quad y_i\left(w^T x_i + b\right) \geq 1, i = 1,2, \dots \dots, n$

③ Dual feasibility $\quad \alpha_i \geq 0, i = 1,2, \dots \dots, n$

④ Complementary slackness $\quad \alpha_i\left(y_i\left(w^T x_i + b\right) - 1\right) = 0$

4번을 이용해서 구해본다.

1) $\alpha_i > 0$ 일 때, 뒤에가 0

⇒ +/- plane위에 위치, 이를 support vector 라고 부른다.

2) $\alpha_i = 0$ 일 때, 뒤에가 0일 필요 없다.

W, b 구하기

목표였던 W를 구하자면

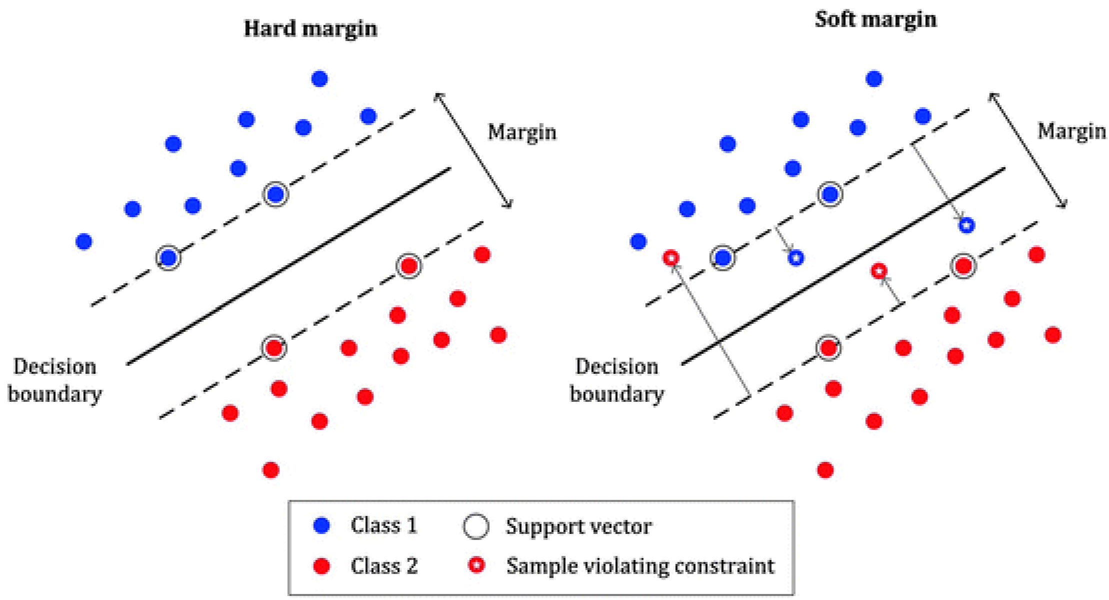
support vector일 때만 $\alpha_i^* \geq 0$ 이니까

support vector만 가지고 decision boundary 구할 수 있다.

$w = \sum_{i=1}^{n} \alpha_i y_i x_i$

b는 support vector 값 하나 (x_{sv}, y_{sv}) 를 넣어봐서 구할 수 있다.

지금까지 살펴본 linear한 SVM을 HardMargin이라고 한다.



Non-linear SVM

soft margin이란

직선으로는 분류하기 힘들다

→ 약간의 error를 허용하자!

목적식 : $minimize \frac{||w||_2^2}{2} + C\sum_{i=1}^{n} \xi_i$

(ξ : 허용하는 error)

Hard margin과 비교해보면 뒤에 정규화 식이 붙었다.

trade-off | C

허용하는 error vs margin

C : trade-off를 결정하는 parameter

C가 높을 수록, error 많이 허용하지 않는다. (하드마진에 가까워진다.)

C가 작을 수록 error 많이 허용한다. (소프트마진에 가까워진다.)

Kernel

SVM에서 선형으로 분리할 수 없는 점들을 분류하기 위해 사용한다.

$f(x) = \sum_{i \in SV} \alpha_i^* y_i K + b$

- Linear Kernel
  - $K < x_1, x_2 >=< x_1, x_2 >$
- Polynomial Kernel
  - $K < x_1, x_2 >= (a < x_1, x_2 > +b)^d$
- Sigmoid Kernel
  - $K < x_1, x_2 >= tanh(a < x_1, x_2 > +b)$
- Gaussian Kernel (rbf kernel)
  - $K < x_1, x_2 >= exp(\frac{||x_1 - x_2||_2^2}{2\sigma^2})$