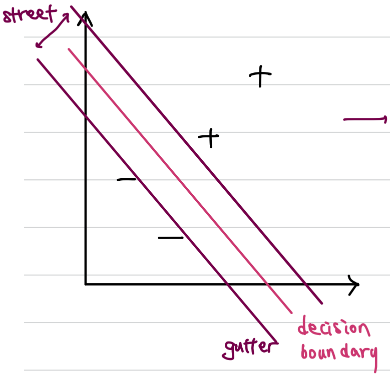
SVM (support vector machines)

유튜브의 MIT OpenCourseWare 16. Learning : Support Vector Machines 강의를 정리

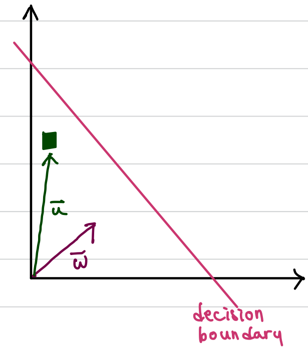
'어떻게 한 공간을 decision boundary를 기준으로 나눌 것인가'에 대한 질문



→ widest street approach 방법

→ street의 넓이가 최대가 되도록

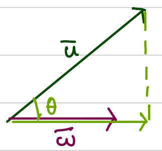
먼저, 해당 decision boundary에 대한 decision rule을 생각해보자.



decision boundary와 직교하는 $\vec{w}$

임의의 점 u를 나타내는 $\vec{u}$ 가 있을 때

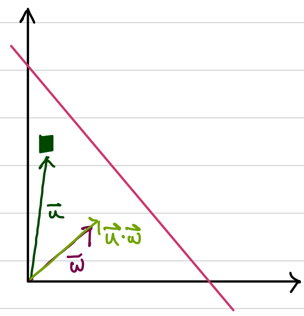
u가 어떤 공간에 속하는지 궁금하다고 하자.



$|\vec{u}||\vec{w}|\cos\theta = \vec{w}\cdot\vec{u}$

$\vec{u}$를 $\vec{w}$ 방향으로 정사영시키면

$\vec{u}$의 decision boundary의 수직 방향으로의 거리를 알 수 있다



∴ decision rule은 다음과 같다

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

$\Leftrightarrow \boxed{\vec{u}\cdot\vec{w} + b \geq 0 \quad \text{then} \quad u \Rightarrow +}$ $(b = -c)$ …①

위 식을 이용하여 decision boundary와 gutter의 거리가 1이라고 한다면, (임의로!)

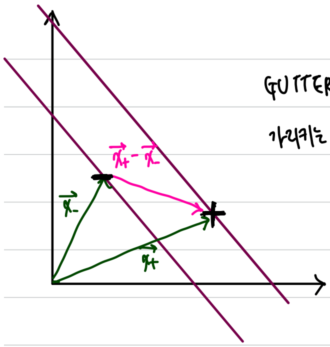
$\vec{w}\cdot\vec{x}_+ + b \geq 1$

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

이때, y_i such that $\begin{cases} y_i = +1 & \text{FOR + SAMPLE} \\ y_i = -1 & \text{FOR − SAMPLE} \end{cases}$ 라고 하면,

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

$\therefore \boxed{y_i(\vec{w}\cdot\vec{x}_i + b) = 1 \quad \text{FOR } x_i \text{ IN GUTTER}}$ …②



GUTTER 위에 존재하는 +SAMPLE, − SAMPLE을 가리키는 벡터를 각각 $\vec{x}_+$, $\vec{x}_-$ 라고 하면,

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

이 때 ②에서

$\begin{cases} \vec{w}\cdot\vec{x}_+ = 1 - b & \text{FOR + SAMPLE} \\ \vec{w}\cdot\vec{x}_- = -1 - b & \text{FOR − SAMPLE} \end{cases}$

$\boxed{\therefore \text{WIDTH} = \dfrac{2}{\|\vec{w}\|}}$ …③

③에서 street width $= \dfrac{2}{\|\vec{w}\|}$ 임을 알았고,

widest street approach에 따라 MAX $\dfrac{2}{\|\vec{w}\|}$ 가 목적이 된다.

$\text{MAX}\dfrac{2}{\|\vec{w}\|} \rightsquigarrow \text{MIN}\|\vec{w}\| \rightsquigarrow \text{MIN}\frac{1}{2}\|\vec{w}\|^2$

MIN $\frac{1}{2}\|\vec{w}\|^2$ 가 되는 $\vec{w}$를 찾기위해 라그랑주 승수법을 사용

> 라그랑주 승수법 (Lagrange Multiplier Method)
> : 제약조건이 있는 최적화문제를 풀기위해 고안된 방법.
> : 최적해의 필요조건을 찾는 방법

$$L = \frac{1}{2}\|\vec{w}\|^2 - \sum \alpha_i \left[y_i(\vec{w}\cdot\vec{x}_i + b) - 1 \right] \quad \cdots ⑦$$

⑦식을 w로 편미분하면

$$\frac{\partial L}{\partial \vec{w}} = \vec{w} - \sum \alpha_i \cdot y_i \cdot \vec{x}_i = 0$$

$$\therefore \boxed{\vec{w} = \sum \alpha_i \cdot y_i \cdot \vec{x}_i} \quad \cdots ④$$

즉, $\vec{w}$는 일부 SAMPLE들의 선형결합으로 나타내어짐

이번엔 ⑦식을 b로 편미분하면

$$\frac{\partial L}{\partial b} = -\sum \alpha_i \cdot y_i = 0$$

$$\therefore \boxed{\sum \alpha_i \cdot y_i = 0} \quad \cdots ⑤$$

⑦식에 ④를 대입해보자

$$L = \frac{1}{2}\left(\sum \alpha_i y_i \vec{x}_i\right)\left(\sum \alpha_j y_j \vec{x}_j\right) - \sum \alpha_i \cdot y_i \cdot \vec{x}_i \cdot \left(\sum \alpha_j \cdot y_j \cdot \vec{x}_j\right) - \overbrace{\sum \alpha_i \cdot y_i}^{} b + \sum \alpha_i$$

상수

$$= \boxed{-\frac{1}{2}\sum\sum \alpha_i \alpha_j y_i y_j \boxed{\vec{x}_i\cdot\vec{x}_j} + \sum \alpha_i} \quad \cdots ⑥$$

↳ 한쌍의 SAMPLE의 내적에 영향을 받음

라그랑주 함수가 α에 대해 2차함수 이므로

$\nabla L = 0$ 이 되는 지점은 하나뿐이다

∴ local minima가 존재하지 않고, 구한 해는 항상 최적의 해임!

다시 decision rule인 ①로 돌아가서 ④를 대입해보면

$$\sum \alpha_i \cdot y_i \cdot \vec{x}_i \bullet \vec{u} + b \geq 0 \quad \text{THEN + SAMPLE}$$

↳ SAMPLE과의 내적과 연관됨

이 $\vec{x}_i$는 위에서 GUTTER 위에 존재하는 점이라고 했고, ($\alpha_i \neq 0$)

이런 SAMPLE이 경계선을 정하기때문에 SVM에서는 이런 샘플을 "support vector" 라고 한다!

경계선에 있지않은 SAMPLE의 모든 $\alpha_i = 0$ 이 된다고 함.

∴ 2개 이상의 SAMPLE이 있으면 widest street approach 방법으로 공간을 (선형적으로) 잘 분류할 수 있음.

Kernel trick

sample을 선형적으로 공간분리를 하기어려운 문제점이 발생.

선형적분리 어려움 　　　　　　　　 공간분리 가능

$\vec{x}$을 공간변환해주는 함수를 ϕ라고하자.

⑥에서 L은 $\vec{x_i} \cdot \vec{x_j}$에 의해 결정되었으므로

결국 관심있는것은 $\phi(\vec{x_i}) \cdot \phi(\vec{x_j})$

$\therefore$ $K(x_i, x_j) = \phi(\vec{x_i}) \cdot \phi(\vec{x_j})$ 로 정의되는 커널함수만 필요함(ϕ이아니라)

커널함수로 사용되는 함수는

① $(\vec{u} \cdot \vec{v} + 1)^n$ 　　 ② $e^{-\frac{\|\vec{u}-\vec{v}\|}{\sigma}}$