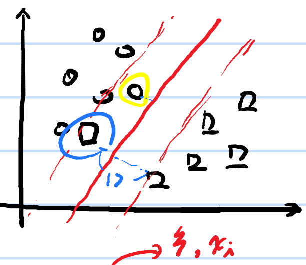
Soft margin

- 선형 분리가 불가능한 데이터셋이거나 Support vector margin 내부에 Datapoint를 허용하는 경우



Slack variable을 사용하는 SVM

- Lagrange multiplier와 비슷한 방식으로, 새로운 변수를 도입.
  - 정상 sample : $\xi_i = 0$
  - 노란 sample : $0 < \xi_i \le 1$ ⇒ 분류는 잘 되어 있으나, margin 내부에 위치하는 Data sample) Decision boundary와 가까운 정도
  - 파란 sample : $\xi_i > 1$ ⇒ 분류에 실패한 Data sample
  - → 모든 Datapoint의 개수만큼 존재.

SVM 최적화 문제 정식화.

* Cost function : minimize $J(\underline{w}, \xi) = \frac{1}{2}\|\underline{w}\|_2^2 + C\sum_{i=1}^{n}\xi_i$

- 첫번째 항 : margin 최대화.
- 두번째 항 : C = 가중치 (모델변수), 어떤 항을 더 중요하게 여길지. $\xi > 0$ 인 sample의 수를 최소화.

($C\Uparrow$: Margin 안에 들어오는 샘플을 덜 허용
$C\Downarrow$: Margin 안에 들어오는 샘플을 어느정도 허용)

Constraints : $y_i(\underline{w}^T\underline{x} + w_0) \ge 1 - \xi_i$, $\xi_i \ge 0$

⇒ 라그랑주 승수법을 이용하여 최적화

- Cost function

$$L(\underline{\alpha}) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j \underline{x}_i \cdot \underline{x}_j$$

- Constraints

$$\sum_{i=1}^{n}\alpha_i y_i = 0, \quad 0 \le \alpha_i \le C, \quad (i = 1, \cdots, n)$$

(달라진 부분)

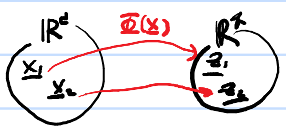
middle

비선형 분포 데이터를 선형 분포로 변환하기.

- 주어진 특징벡터 $\underline{x} \in \mathbb{R}^d$를 변환하여 q차원 vector로 변환할 수 있음. ($q \ge d$)

$$\Phi(\underline{x}) = \Phi((x_1, x_2, \cdots, x_d)^T)$$
$$= (\phi_1(\underline{x}), \phi_2(\underline{x}), \cdots, \phi_q(\underline{x}))^T = \underline{z}$$

- $\Phi(\underline{x})$: 임의의 비선형함수.



($\Phi(\underline{x}) = \mathbb{R}^2 \to \mathbb{R}^1$
$\Phi(x_1, x_2) = x_1 = \underline{z}$) ⇒ 차원 축소.

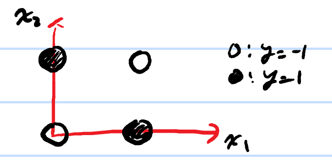
($\Phi(\underline{x}) : \mathbb{R}^2 \to \mathbb{R}^3$
$\Phi(x_1, x_2) = (x_1, x_1^2, x_1^3)^T = \underline{z}$) ⇒ 차원 증식

"분포를 다루기 좋게 변환"

XOR 분류문제

$\underline{x}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, y_1 = -1$ $\quad \underline{x}_3 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, y_3 = 1$

$\underline{x}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, y_2 = 1$ $\quad \underline{x}_4 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, y_4 = -1$



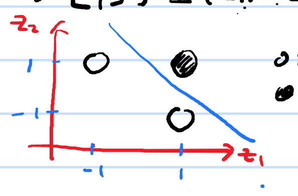
* Perceptron을 사용한 변환 (MLP의 예시)

$$\Phi(\underline{x}) = \begin{bmatrix} \phi_1(\underline{x}) \\ \phi_2(\underline{x}) \end{bmatrix} = \begin{bmatrix} \mathrm{sign}(x_1 + x_2 - 0.5) \\ \mathrm{sign}(-x_1 - x_2 + 1.5) \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \underline{z}$$

$\underline{w} = [1, 1], w_0 = -0.5$

$\underline{w} = [-1, -1], w_0 = 1.5$

- $\Phi(\underline{x}_1) = z_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, $\Phi(\underline{x}_2) = z_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$
  $\Phi(\underline{x}_3) = z_3 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\Phi(\underline{x}_4) = z_4 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$,,



⇒ 변환된 특징벡터 데이터셋 $(\Phi(x_i), y_i) = (z_i, y_i)$는 선형 분리 가능하게 바뀌었다.

- Example

$$\Phi(\underline{x}) = \begin{bmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2}x_1x_2 \end{bmatrix} = \underline{z} \in \mathbb{R}^2$$

$\Phi(\underline{x}_1) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, $\Phi(\underline{x}_2) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\Phi(\underline{x}_3) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, $\Phi(\underline{x}_4) = \begin{bmatrix} 1 \\ 1 \\ \sqrt{2} \end{bmatrix}$