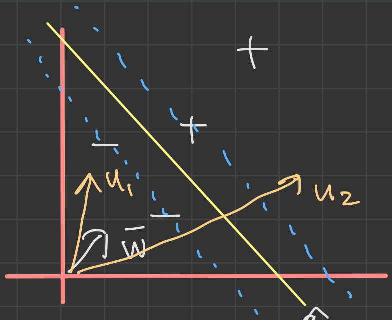


↑ represented by $\bar{w}$

Decision rule

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

distance to the yellow line

length of $\bar{u}$ on direction of $\bar{w}$

$$\bar{w} \cdot \bar{u} + b \geq 0 \quad ①$$

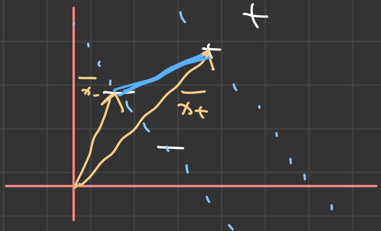
Introduce y_i (label i) $\begin{cases} +1 & \text{for + sample} \\ -1 & \text{for − sample} \end{cases}$

Constraint

$$y_i(\bar{w} \cdot \bar{x}_i + b) - 1 \geq 0 \quad ②$$

make $y_i(\bar{w} \cdot \bar{x}_i + b) - 1 = 0$ for samples on gutter

"- - - - - -"



$$\text{width} = (\overline{x_+} - \overline{x_-}) \cdot \frac{\overline{w}}{\|\overline{w}\|}$$

from 2

$\overline{x_+} \cdot \overline{w} = 1 - b$

$\overline{x_-} \cdot \overline{w} = 1 + b$

so width $= \frac{2}{\|w\|}$ 3

want to maximize

minimize $\frac{1}{2}\|w\|^2$ for mathematical convience.

using lagrange multipliers

$$L = \frac{1}{2}\|w\|^2 - \sum \alpha_i [y_i(\overline{w} \cdot \overline{x_i} + b) - 1]$$

$$\frac{\partial L}{\partial w} = \overline{w} - \sum \alpha_i y_i x_i = 0$$

so $\overline{w} = \sum \alpha_i y_i \overline{x_i}$

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

plug $\overline{w} = \sum \alpha_i y_i \overline{x_i}$ back in L

$$L = \frac{1}{2}\left(\sum \alpha_i y_i \overline{x_i}\right)\left(\sum \alpha_j y_j \overline{x_j}\right) - \sum \alpha_i y_i \overline{x_i} \cdot \sum \alpha_j y_j \overline{x_j}$$

$$- \underbrace{\sum \alpha_i y_i b}_{\substack{=0 \text{ because } \sum \alpha_i y_i = 0 \\ \text{and } b \text{ is a constant}}} - \sum \alpha_i$$

$$L = \sum \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \boxed{x_i \cdot x_j}$$

change to $K(x_i, x_j)$ when non-linear.

4

from ① and ④

decision rule:

$$\underbrace{\sum \alpha_i y_i \bar{x}_i}_{\bar{w}} \cdot \bar{u} + b \geq 0 \quad \text{THEN } +$$

④ on book:

$$\max_{\alpha} \left[\sum_{i=1}^{m} \alpha - \frac{1}{2} \sum_{i,j=1}^{m} label^{(i)} \cdot label^{(j)} \cdot a_i \cdot a_j \langle x^{(i)}, x^{(j)} \rangle \right]$$

subject to the following constraints:

$label^{(i)} \equiv y_i$

$$\boxed{\alpha \geq 0,} and \sum_{i-1}^{m} \alpha_i \cdot label^{(i)} = 0$$

$$\boxed{c \geq \alpha \geq 0,} and \sum_{i-1}^{m} \alpha_i \cdot label^{(i)} = 0$$

The constant C controls weighting between our goal of making the margin large and ensuring that most of the examples have a functional margin of at least 1.0. The constant C is an argument to our optimization code that we can tune and get different results. Once we solve for our alphas, we can write the separating hyperplane in terms of these alphas. That part is straightforward. The majority of the work in SVMs is finding the alphas.

SMO

Sequential Minimal Optimization

Idea: update 2 alphas each time and maintain $\sum \alpha_i y_i = 0$

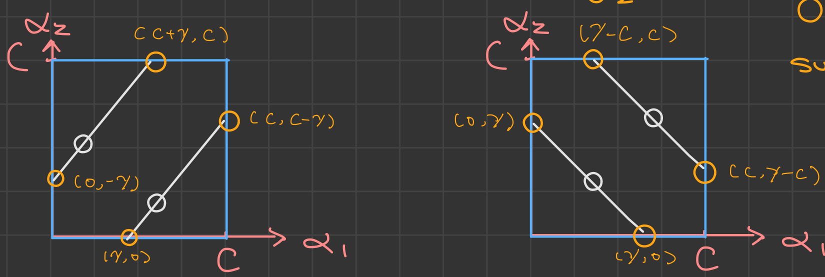
α' : updated α

$$\alpha_1 y_1 + \alpha_2 y_2 = \alpha_1' y_1 + \alpha_2' y_2 = \text{constant}.$$

two situation

let $\gamma = \alpha_1 + y_1 y_2 \alpha_2$

O : bounds for α_1 and α_2 such that $0 \le \alpha_1, \alpha_2 \le C$



(α_1, α_2) can only move on the line

$$\alpha_1 y_1 + \alpha_2 y_2 = \text{Constant}$$

Clip : $\alpha = \max(\alpha, U)$
$\alpha = \min(\alpha, L)$

U : upper bound
L : lower bound.

Update function

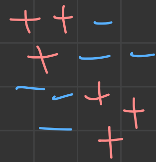
$$\alpha_2' = \text{Clip}\left(\alpha_2 + \frac{y_2 \cdot (E_2 - E_1)}{2K_{12} - K_{11} - K_{22}} \right)$$

where $E_i = \underbrace{\left(\sum_{j=1}^{m} \alpha_j y_j K_{ji} + b \right)}_{\text{ESTIMATE } f(x_i)} - \underbrace{y_i}_{\text{LABEL}}$

E_i : ERROR

Kernels:

To find a "hyperplane" to seperate the data instead of a line, we use kernels instead of dot product for k_{ij}



radial bias function:

$$K(x, y) = \exp\left(\frac{-\|x-y\|^2}{2\sigma^2}\right)$$

σ : a variable that can be tuned

$\langle x-y, x-y \rangle$

Decision rule when using Kernels:

$$\sum \alpha_i y_i \, K(\bar{x}_i, \bar{u}) + b > 0$$

dot product for linear / non-kernel