SVM

January 22, 2022

Contents

1 Introduction

Sample Space:

$$D = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\} \tag{1}$$

Hyperplane can split Sample Space into two partitions:

$$\boldsymbol{\omega}^T \boldsymbol{x} + b = 0 \tag{2}$$

The distance of each point $\boldsymbol{x}_i$ from Hyperplane $(\boldsymbol{\omega}, b)$ can be written as:

$$r = \frac{|\boldsymbol{\omega}^T \boldsymbol{x} + b|}{\|\boldsymbol{\omega}\|} \tag{3}$$

The Decision Rule:

$$r = \frac{|\boldsymbol{\omega}^T \boldsymbol{x} + b|}{\|\boldsymbol{\omega}\|} \geq 0 \Rightarrow \boldsymbol{x} \; is \; positive \; sample \tag{4}$$

We can use factor to scale Hyperplane to $(\alpha\boldsymbol{\omega}, \beta b)$ so that the rule of seperate two partitions can be written as (For Mathematical Convenient):

$$\begin{cases} \boldsymbol{\omega}^T \boldsymbol{x}_i + b \geq +1 & y_i = +1 \\ \boldsymbol{\omega}^T \boldsymbol{x}_i + b \leq -1 & y_i = -1 \end{cases} \tag{5}$$

The distance of Surfuce Hyperplanes $(\boldsymbol{\omega}, b+1)$ from $(\boldsymbol{\omega}, b-1)$

$$\gamma = \frac{2}{\|\omega\|} \tag{6}$$

Our destination is to maximize the distance in the constrains 5

$$\begin{gathered} \max_{\boldsymbol{\omega}, b} \frac{2}{\|\omega\|} \\ y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) \geq 1 \end{gathered} \tag{7}$$

The Support Vectors satisify:

$$y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) - 1 = 0 \tag{8}$$

It's akin to minimize the norm of ω

$$\begin{gathered} \min_{\boldsymbol{\omega}, b} \frac{\|\omega\|^2}{2} \\ y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) \geq 1 \end{gathered} \tag{9}$$

2 Dual

Using Lagrange Multiplier Method

$$L(\boldsymbol{\omega}, b, \boldsymbol{\alpha}) = \frac{\|\boldsymbol{\omega}\|^2}{2} + \sum_{i=1}^{n} \alpha_i [1 - y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b)] \tag{10}$$

$$\begin{aligned} \frac{\partial L}{\partial \boldsymbol{\omega}} &= \boldsymbol{\omega} - \sum_{i=1}^{n} \alpha_i y_i \boldsymbol{x}_i = 0 \\ \frac{\partial L}{\partial b} &= -\sum_{i=1}^{n} \alpha_i y_i = 0 \end{aligned} \tag{11}$$

$$L(\boldsymbol{\omega}, b, \boldsymbol{\alpha}) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j \tag{12}$$

The Dual Form of SVM:

$$\begin{aligned} &\min_{\boldsymbol{\alpha}} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j \\ &\sum_{i=1}^{n} \alpha_i y_i = 0 \\ &\boldsymbol{\alpha} \geq 0 \end{aligned} \tag{13}$$

$$f(\boldsymbol{x}) = \boldsymbol{\omega}^T \boldsymbol{x} + b = \sum_{i}^{n} \alpha_i y_i \boldsymbol{x}_i^T \boldsymbol{x} + b \tag{14}$$

It meets the KKT rules

$$\begin{aligned} &\boldsymbol{\alpha} \geq 0 \\ &y_i f(\boldsymbol{x}_i) - 1 \geq 0 \\ &\alpha_i (y_i f(\boldsymbol{x}_i) - 1) = 0 \end{aligned} \tag{15}$$

When $\alpha_i = 0$, $(\boldsymbol{x}_i, y_i)$ will be ignored
When $\alpha_i \geq 0$, $y_i f(\boldsymbol{x}_i) = 1$ which means $(\boldsymbol{x}_i, y_i)$ in the Surfuce Hyper-planes.

3 SMO

Select two variable α_i,α_j and fix other variable

$$\begin{aligned}
&\min_{\alpha_i,\alpha_j} \alpha_i + \alpha_j - \frac{1}{2}\alpha_i^2 y_i^2 \|\boldsymbol{x}_i\|^2 - \alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j - \frac{1}{2}\alpha_j^2 y_j^2 \|\boldsymbol{x}_j\|^2 \\
&- \sum_{k\neq i,j} \alpha_i \alpha_k y_i y_k \boldsymbol{x}_i^T \boldsymbol{x}_k - \sum_{k\neq i,j} \alpha_j \alpha_k y_j y_k \boldsymbol{x}_j^T \boldsymbol{x}_k \\
&\alpha_i y_i + \alpha_j y_j = \sum_{k\neq i,j} \alpha_k y_k = \varsigma \\
&\alpha_i \geq 0, \alpha_j \geq 0
\end{aligned} \tag{16}$$

$$a_j = \varsigma y_j - \alpha_i y_i y_j \tag{17}$$

Solve the equation

$$\begin{aligned}
&(\|\boldsymbol{x}_i\|^2 - 2\boldsymbol{x}_i^T \boldsymbol{x}_j + \|\boldsymbol{x}_j\|^2)\alpha_i \\
&= 1 - y_i y_j - \varsigma y_i \boldsymbol{x}_i^T \boldsymbol{x}_j + \varsigma y_i \|\boldsymbol{x}_j\|^2 \\
&- y_i \sum_{k\neq i,j} \alpha_k y_k \boldsymbol{x}_i^T \boldsymbol{x}_k + y_i \sum_{k\neq i,j} \alpha_k y_k \boldsymbol{x}_j^T \boldsymbol{x}_k
\end{aligned} \tag{18}$$

Owing to the complexity, We want to replace ς

$$\begin{aligned}
f(\boldsymbol{x}_i) &= \sum_{k}^{n} \alpha_k y_y \boldsymbol{x}_k^T \boldsymbol{x_i} + b \\
&= \alpha_i y_i \|\boldsymbol{x}_i\|^2 + \alpha_j y_j \boldsymbol{x}_j^T \boldsymbol{x_i} + \sum_{k\neq i,j} \alpha_k y_k \boldsymbol{x}_i^T \boldsymbol{x}_k + b \\
f(\boldsymbol{x}_j) &= \sum_{k}^{n} \alpha_k y_y \boldsymbol{x}_k^T \boldsymbol{x_j} + b \\
&= \alpha_j y_j \|\boldsymbol{x}_j\|^2 + \alpha_i y_i \boldsymbol{x}_i^T \boldsymbol{x_j} + \sum_{k\neq i,j} \alpha_k y_k \boldsymbol{x}_j^T \boldsymbol{x}_k + b
\end{aligned} \tag{19}$$

$$\begin{aligned}
&\sum_{k\neq i,j} \alpha_k y_k \boldsymbol{x}_j^T \boldsymbol{x}_k - \sum_{k\neq i,j} \alpha_k y_k \boldsymbol{x}_i^T \boldsymbol{x}_k \\
&= f(\boldsymbol{x}_j) - f(\boldsymbol{x}_i) \\
&+ \alpha_i y_i \|\boldsymbol{x}_i\|^2 - \alpha_j y_j \|\boldsymbol{x}_j\|^2 \\
&+ \alpha_j y_j \boldsymbol{x}_j^T \boldsymbol{x_i} - \alpha_i y_i \boldsymbol{x}_i^T \boldsymbol{x_j}
\end{aligned} \tag{20}$$

$$
\begin{aligned}
&y_i\left[\sum_{k\neq i,j}\alpha_k y_k \boldsymbol{x}_j^T\boldsymbol{x}_k-\sum_{k\neq i,j}\alpha_k y_k \boldsymbol{x}_i^T\boldsymbol{x}_k\right]\\
&=y_i\left(f(\boldsymbol{x}_j)-f(\boldsymbol{x}_i)\right)\\
&+\alpha_i\|\boldsymbol{x}_i\|^2-(\varsigma y_j-\alpha_i y_i y_j)y_i y_j\|\boldsymbol{x}_j\|^2\\
&+(\varsigma y_j-\alpha_i y_i y_j)y_i y_j\boldsymbol{x}_j^T\boldsymbol{x_i}-\alpha_i\boldsymbol{x}_i^T\boldsymbol{x_j}\\
&=y_i\left(f(\boldsymbol{x}_j)-f(\boldsymbol{x}_i)\right)\\
&+\left(\|\boldsymbol{x}_i\|^2-2\boldsymbol{x}_i^T\boldsymbol{x_j}+\|\boldsymbol{x}_j\|^2\right)\alpha_i\\
&+\varsigma y_i\boldsymbol{x}_i^T\boldsymbol{x_j}-\varsigma y_i\|\boldsymbol{x}_j\|^2
\end{aligned}
\tag{21}
$$

$$
\begin{aligned}
&(\|\boldsymbol{x}_i\|^2-2\boldsymbol{x}_i^T\boldsymbol{x}_j+\|\boldsymbol{x}_j\|^2)\alpha_i^*\\
&=1-y_i y_j-\varsigma y_i\boldsymbol{x}_i^T\boldsymbol{x}_j+\varsigma y_i\|\boldsymbol{x}_j\|^2\\
&+y_i\left(f(\boldsymbol{x}_j)-f(\boldsymbol{x}_i)\right)\\
&+\left(\|\boldsymbol{x}_i\|^2-2\boldsymbol{x}_i^T\boldsymbol{x_j}+\|\boldsymbol{x}_j\|^2\right)\alpha_i\\
&+\varsigma y_i\boldsymbol{x}_i^T\boldsymbol{x_j}-\varsigma y_i\|\boldsymbol{x}_j\|^2\\
&=\left(\|\boldsymbol{x}_i\|^2-2\boldsymbol{x}_i^T\boldsymbol{x_j}+\|\boldsymbol{x}_j\|^2\right)\alpha_i\\
&+1-y_i y_j+y_i\left(f(\boldsymbol{x}_j)-f(\boldsymbol{x}_i)\right)
\end{aligned}
\tag{22}
$$

$$
\begin{aligned}
\alpha_i^*&=\alpha_i+\frac{y_i\left[(f(\boldsymbol{x}_j)-y_j)-(f(\boldsymbol{x}_i)-y_i)\right]}{\|\boldsymbol{x}_i\|^2-2\boldsymbol{x}_i^T\boldsymbol{x}_j+\|\boldsymbol{x}_j\|^2}\\
\alpha_j^*&=\alpha_j+\frac{y_j\left[(f(\boldsymbol{x}_i)-y_i)-(f(\boldsymbol{x}_j)-y_j)\right]}{\|\boldsymbol{x}_i\|^2-2\boldsymbol{x}_i^T\boldsymbol{x}_j+\|\boldsymbol{x}_j\|^2}
\end{aligned}
\tag{23}
$$

Consider the range of α_i,α_j

$$
\begin{aligned}
\alpha_i\in[0,C]\\
\alpha_j\in[0,C]
\end{aligned}
\tag{24}
$$

When $y_i=y_j$,$\alpha_i+\alpha_j=\varsigma y_i$

$$
\begin{aligned}
\inf\alpha_i^*&=\max\{0,\alpha_j+\alpha_i-C\}\\
\sup\alpha_i^*&=\min\{C,\alpha_j+\alpha_i\}
\end{aligned}
\tag{25}
$$

When $y_i\neq y_j$,$\alpha_i-\alpha_j=\varsigma y_i$

$$
\begin{aligned}
\inf \alpha_i^* &= \max\{0, \alpha_j - \alpha_i\} \\
\sup \alpha_i^* &= \min\{C, \alpha_j - \alpha_i + C\}
\end{aligned}
\tag{26}
$$

$$
\begin{aligned}
\bar{\alpha}_i &= \begin{cases} \sup \alpha_i^*, \alpha_i^* \geq \sup \alpha_i^* \\ \alpha_i^*, \inf \alpha_i^* \leq \alpha_i^* \leq \sup \alpha_i^* \\ \inf \alpha_i^*, \alpha_i^* \leq \inf \alpha_i^* \end{cases} \\
\bar{\alpha}_j &= \alpha_j + (\alpha_i - \bar{\alpha}_i) y_i y_j
\end{aligned}
\tag{27}
$$

After the computation of α_i,α_j, we need to calculate b
We maintain the Support Vector Set S in the Surface Hyperplanes

$$
\begin{aligned}
(\boldsymbol{x}_s, y_s) &\in S \\
y_s(\boldsymbol{\omega}^T \boldsymbol{x}_s + b) &= 1 \\
\boldsymbol{\omega} &= \sum_{i \in S} \alpha_i y_i \boldsymbol{x}_i^T
\end{aligned}
\tag{28}
$$

$$
b + \sum_{i \in S} \alpha_i y_i \boldsymbol{x}_i^T \boldsymbol{x}_s = y_s \tag{29}
$$

$$
\bar{b} = \frac{1}{|S|} \sum_{s \in S} \left(y_s - \sum_{i \in S} \alpha_i y_i \boldsymbol{x}_i^T \boldsymbol{x}_s \right) \tag{30}
$$

Also

$$
\bar{b}_i = y_i - \bar{\alpha}_i y_i \|\boldsymbol{x}_i\|^2 - \bar{\alpha}_j y_j \boldsymbol{x}_i^T \boldsymbol{x}_j - \sum_{k \neq i,j}^{n} \alpha_k y_k \boldsymbol{x}_k^T \boldsymbol{x}_i \tag{31}
$$

$$
\sum_{k \neq i,j}^{n} \alpha_k y_k \boldsymbol{x}_k^T \boldsymbol{x}_i = f(\boldsymbol{x}_i) - \alpha_i y_i \|\boldsymbol{x}_i\|^2 - \alpha_j y_j \boldsymbol{x}_i^T \boldsymbol{x}_j - b \tag{32}
$$

$$
\begin{aligned}
\bar{b}_i &= b - (f(\boldsymbol{x}_i) - y_i) + (\alpha_i - \bar{\alpha}_i) y_i \|\boldsymbol{x}_i\|^2 + (\alpha_j - \bar{\alpha}_j) y_j \boldsymbol{x}_i^T \boldsymbol{x}_j \\
\bar{b}_j &= b - (f(\boldsymbol{x}_j) - y_j) + (\alpha_j - \bar{\alpha}_j) y_j \|\boldsymbol{x}_j\|^2 + (\alpha_i - \bar{\alpha}_i) y_i \boldsymbol{x}_i^T \boldsymbol{x}_j
\end{aligned}
\tag{33}
$$

$$
\bar{b} = \frac{\bar{b}_i + \bar{b}_j}{2} \tag{34}
$$

4 Kernel Function

Transform Function $\phi(\boldsymbol{x})$ map the vector $\boldsymbol{x}$ into other space

$$\begin{aligned}
&\min_{\boldsymbol{\alpha}} \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j y_i y_j \phi(\boldsymbol{x}_i)^T \phi(\boldsymbol{x}_j) \\
&\sum_{i=1}^{n} \alpha_i y_i = 0 \\
&\boldsymbol{\alpha} \geq 0
\end{aligned} \tag{35}$$

We can construct the Kernel Function instead of knowning the Transform Function

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi(\boldsymbol{x}_i)^T \phi(\boldsymbol{x}_j) \tag{36}$$

Kernel Function satisify Mercer Theory

$$\kappa = \{K(\boldsymbol{x}_i, \boldsymbol{x}_j)\} \succeq 0 \tag{37}$$

Polynomial Kernel

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i \boldsymbol{x}_j + 1)^n \tag{38}$$

Gaussian Kernel

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = e^{-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{2\sigma^2}} \tag{39}$$

Laplacian Kernel

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = e^{-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|}{\sigma}} \tag{40}$$

Sigmoid Kernel

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \tanh(\beta \boldsymbol{x}_i^T \boldsymbol{x}_j + \theta) \tag{41}$$

$$\begin{aligned}
&\gamma_1 K_1 + \gamma_2 K_2 \\
&K_1 \otimes K_2(\boldsymbol{x}_i, \boldsymbol{x}_j) = K_1(\boldsymbol{x}_i, \boldsymbol{x}_j) K_2(\boldsymbol{x}_i, \boldsymbol{x}_j) \\
&f(\boldsymbol{x}_i) K(\boldsymbol{x}_i, \boldsymbol{x}_j) f(\boldsymbol{x}_j)
\end{aligned} \tag{42}$$

5 RKHS