HW8. Support Vector Machines

Name - Sayem Lincoln

PID - A54207835

1) We have $Q = \begin{bmatrix} 0 & 0_d^T \\ 0_d & I_d \end{bmatrix}$ where $\begin{cases} 0_d^T = [0 \cdots\cdots 0] \in \mathbb{R}^d \\ I_d = d \times d \text{ identity matrix} \end{cases}$

We have to prove that Q is a positive semi-definite matrix

Q positive semidefitie $\Leftrightarrow u^T Q u \geq 0$ for any u

for $u \in \mathbb{R}^{d+1} \Rightarrow u = \begin{bmatrix} u_1 \\ \vdots \\ u_{d+1} \end{bmatrix}$

$$u^T Q u = \sum_{i=2}^{d+1} u_i^2 \geq 0$$

$\Rightarrow$ Thus Q is positive semi-definite.

$$Q = \begin{pmatrix} 0 & \overbrace{(0 \cdots\cdots 0)}^{d} \\ \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} & \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} \end{pmatrix}$$

2) a) minimize $\frac{1}{2} \sum_{m=1}^{N} \sum_{n=1}^{N} y_n y_m \alpha_n \alpha_m x_n^T x_m - \sum_{n=1}^{N} \alpha_n$ (i)

subject to $\sum_{n=1}^{N} y_n \alpha_n = 0$; $\alpha_n \geq 0$ $(n = 1, \ldots, N)$ (ii)

Problem written as $\Rightarrow$

minimize $\Rightarrow \frac{1}{2} u^T Q u + p^T u$

for any u subject to $\Rightarrow a_m^T u \geq c_m$ $(m = 1, \ldots, N)$

we begin by simplifying the expression of (i) $\Rightarrow$

$$\frac{1}{2} \sum_{m=1}^{N} \sum_{n=1}^{N} y_n y_m \alpha_n \alpha_m x_N^T x_M - \sum_{n=1}^{N} \alpha_m =$$

$$= \frac{1}{2} [\alpha_1 \cdots \alpha_N] \underbrace{\begin{bmatrix} y_1 y_1 x_1^T x_1 & \cdots & y_1 y_N x_1^T x_N \\ y_2 y_1 x_2^T x_1 & \cdots & y_2 y_N x_2^T x_N \\ \vdots & & \vdots \\ y_N y_1 x_N^T x_1 & \cdots & y_N y_N x_N^T x_N \end{bmatrix}}_{D} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} -$$

$$[1 \cdots 1] \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix}$$

$$= \frac{1}{2} \alpha^T Q_D \alpha - 1_N^T \alpha$$

We Simplify the expression (ii) $\Rightarrow$

we have $\sum_{m=1}^{N} y_n \alpha_n = 0 \Rightarrow \begin{cases} \sum_{m=1}^{N} y_n \alpha_n \geq 0 \\ \sum_{m=1}^{N} y_m \alpha_n \leq 0 \end{cases}$

$$\Rightarrow \begin{cases} \sum_{m=n}^{N} y_n \alpha_n \geq 0 \\ \sum_{m=n}^{N} -y_n \alpha_n \geq 0 \end{cases}$$

we have also $\alpha_m \geq 0$ $(m = 1, \ldots, N)$

$$\Rightarrow \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} \geq O_N$$

Thus we can write $\Rightarrow$

$$\begin{bmatrix} [y_1 \ldots y_N] \\ [-y_1 \ldots -y_N] \\ \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} \geq O_{N+2}$$

$$\Rightarrow A_D \times \alpha \geq O_{N+2}$$

Finally, we conclude that the problem is a Standard QP- Problem $\Rightarrow$

$$\begin{cases} \min_{\alpha \in \mathbb{R}} \frac{1}{2} \alpha^T Q_D \alpha - I_N^T \alpha \end{cases}$$

$$Q_D = \begin{bmatrix} y_1 y_1 x_1^T x_1 & \cdots & y_1 y_N x_1^T x_N \\ \vdots & & \vdots \\ y_N y_1 x_N^T x_1 & \cdots & y_N y_N x_N^T x_N \end{bmatrix} \quad \text{and} \quad A_D = \begin{bmatrix} y^T \\ -y^T \\ I_{N \times N} \end{bmatrix}$$

b) We suppose ⇒

$$X_s = \begin{bmatrix} -y_1 x_1^T \\ | \\ -y_N x_N^T \end{bmatrix}$$

$$X_s X_s^T = \begin{bmatrix} -y_1 x_1^T \\ | \\ -y_N x_N^T \end{bmatrix} \begin{bmatrix} -y_1 x_1^T & \cdots & -y_N x_N^T \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} y_1 y_1 x_1^T x_1 & \cdots & y_1 y_N x_1^T x_N \\ | & & | \\ y_N y_1 x_N^T x_N & \cdots & y_N y_N x_N^T x_N \end{bmatrix}}_{Q_D}$$

$$= Q_D$$

$$\Rightarrow Q_D = X_s X_s^T$$

Q_D is positive semi-definite $\iff u^T Q_D u \geq 0$ for any u

$$\iff u^T X_s X_s^T u \geq 0$$

$$\iff (u^T X_s)(u^T X_s)^T \geq 0$$

$$\iff \| u^T X_s \|^2 \geq 0$$

Thus, Q_D is positive semi-definite.

3) Assume D (data set) with two data points $(x_+, +1)$ and $(x_-, -1)$.
We compute hyper plane and its margin (b^*, w^*).
The two seperation constraints are ⇒

$$\begin{cases} (w^T x_+ + b) \geq 1 & — \text{(i)} \\ \text{and} \\ -(w^T x_- + b) \geq 1 & — \text{(ii)} \end{cases}$$

By adding (i) and (ii) $\Rightarrow w^T(x_+ - x_-) \geq 2$

The Cauchy - Schwarz inequality gives us ⇒

$$\Rightarrow 2 \leq w^T(x_+ - x_-) \leq \left| w^T(x_+ - x_-) \right| \leq \|w\| \cdot \|x_+ - x_-\|$$

So we get $\Rightarrow \|w\| \geqslant \dfrac{2}{\|x_+ - x_-\|}$

Thus $\Rightarrow \|w^*\| = \dfrac{2}{\|x_+ - x_-\|}$ (The minimum)

In this case we want w^* to satisfy both constraints (1) & (4). This means that $\Rightarrow$

$$\left| w^{*T}(x_+ - x_-) \right| = \|w^*\| \cdot \|x_+ - x_-\|$$

This can happen only if $\Rightarrow$

$$w^* = k(x_+ - x_-)$$

but we have $\|w^*\| = \dfrac{2}{\|x_+ - x_-\|}$

$$\Rightarrow k = \frac{2}{\|x_+ - x_-\|^2}$$

Finally we may write $\Rightarrow w^* = \dfrac{2(x_+ - x_-)}{\|x_+ - x_-\|^2}$

It remains to be determined $b^* \Rightarrow$

we have $\Rightarrow 2\left(\dfrac{(x_+ - x_-)}{\|x_+ - x_-\|^2}\right)^T x_+ + b^* = 1$

$$\Rightarrow b^* = 1 - 2\,\frac{x_+^T x_+ - x_-^T x_+}{\|x_+ - x_-\|^2}$$

$$= \frac{\|x_-\|^2 - \|x_+\|^2}{\|x_+ - x_-\|^2}$$

Thus, $b^* = \dfrac{\|x_-\|^2 - \|x_+\|^2}{\|x_+ - x_-\|^2}$

Finally, we can say that (w^*, b^*) satisfies both constraints and minimizes $\|w\|$, and therefore gives us the optimal hyperplane.

4) We have a data set of three data points in $\mathbb{R}^2 \Rightarrow$

$$X = \begin{pmatrix} 0 & 0 \\ 0 & -1 \\ -2 & 0 \end{pmatrix} \qquad y = \begin{pmatrix} -1 \\ -1 \\ +1 \end{pmatrix}$$

In this case the constraints are $\Rightarrow$

$$\begin{cases} -b \geq 1 & \text{— (i)} \\ -(w_2 + b) \geq 1 & \text{— (ii)} \\ (-2w_1 + b) \geq 1 & \text{— (iii)} \end{cases}$$

We combine (i) and (ii) $\Rightarrow$ $w_1 \leq -1$

if we consider $W = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$, the quantity we seek to minimize is $\Rightarrow$ $\frac{1}{2} w^T w = \frac{1}{2}(w_1^2 + w_2^2) \geq \frac{1}{2}(1 + 0) \geq \frac{1}{2}$

We have equality when $w_1 = -1$ and $w_2 = 0$. So, when $w^* = (-1, 0)$, the constraint (iii) gives us $\Rightarrow b \geq -1$.

So we choose $b^* = -1$.

Thus, we verify (w^*, b^*) satisfies all constraints and minimizes $\|w\|$ and therefore gives us the optimal hyperplane.

NB the margin in this case is $\frac{1}{\|w^*\|} = 1$.