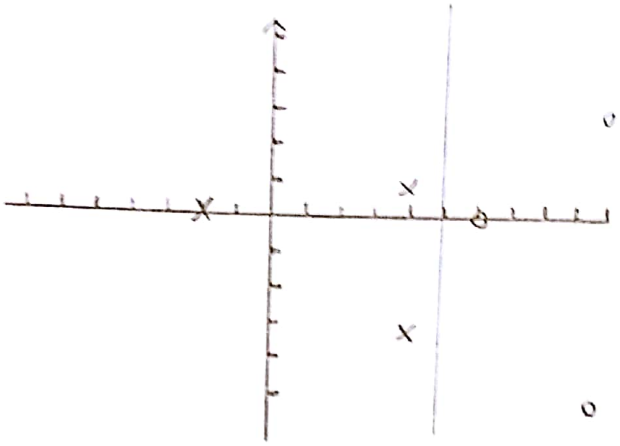
ML Tech HW1

r06922114 陳啟元

1. $z_1 = (-2, 0)$, $z_2 = (4, -3)$, $z_3 = (4, 1)$, $z_4 = (6, 0)$, $z_5 = (10, -5)$,
$z_6 = z_7 = (10, 3)$



Also, by solving primal hard margin SVM with package cvxopt's QP solver we have $(b, w) = (-5, [1, 6.25\times10^{-9}]) \approx (-5, [1, 0])$, which imply the equation $z_1 \cdot 1 + z_2 \cdot 0 - 5 = 0$. It's the same as above.

2. By the package mentioned above, we obtain

$\alpha = (2.39\times10^{-17}, 2.5\times10^{-1}, 2.5\times10^{-1}, 3.33\times10^{-1}, 8.33\times10^{-2}, 8.33\times10^{-2}, 3.89\times10^{-9})$

α_1 and α_7 are close to zero, So we "believe" that $(x_1, y_1), (x_7, y_7)$ are not SVs. $(x_2, y_2) \sim (x_6, y_6)$ are support vectors

3. We know that $w = \sum_{SV} \alpha_n y_n z_n$. $\Phi_k(x) = (1, 2x_1, 2x_2, 2x_1^2, 2x_1x_2, 2x_2x_1, 2x_2^2)$

By the α above, $w = \begin{bmatrix} 4.85\times10^{-9} \\ -6.66\times10^{-1} \\ 2.15\times10^{-14} \\ 6.66\times10^{-1} \\ 0 \\ 0 \\ 3.33\times10^{-1} \end{bmatrix}_{7\times1}$, $B = \begin{bmatrix} -1 \\ -1.66 \\ -1.66 \\ -1.66 \\ -1.66 \\ -1.66 \\ -17 \end{bmatrix}$ the ith element is the b we calculated by considering (x_i, y_i) as SV.

$\boxed{w = \sum_{SV} \alpha_n y_n z_n \approx \sum_{all} \alpha_n y_n z_n \quad \text{So we obtain } w \text{ by } \sum_{all} \alpha_n y_n z_n}$

where $b = y_s - w^T z_s = y_s - \sum_{SV} \alpha_n y_n z_n \cdot z_s = y_s - \sum_{SV} \alpha_n y_n K(x_n, x_s)$

(x_s, y_s) is a support vector.

We can conclude that $(x_2, y_2) \sim (x_6, y_6)$ are SVs ($\because$ They can obtain same b)

The curve in X is: $w^T \Phi_k(x) - 1.66 = 0$

4. problem 1's curve: $\delta_1 - 5 = 0 \Leftrightarrow (2x_2^2 - 4x_1 + 2) - 5 = 0$

problem 3's curve: $W^T \Phi_k(x) - 1.66 = 0$ $\quad 2x_2^2 - 4x_1 - 3 = 0$

$$\Leftrightarrow 4.85\times10^{-9} - 6.66\times10^{-1}\times 2x_1 + 2.15\times10^{-14}\times 2x_2 + 6.66\times10^{-1}\times 2x_1^2 + 3.33\times10^{-1}\times 2x_2^2 - 1.66 = 0$$

$$\approx 6.66\times10^{-1}\times 2x_1 + 6.66\times10^{-1}\times 2x_1^2 + 3.33\times10^{-1}\times 2x_2^2 - 1.66 = 0$$

it has x_1^2 which problem 1's doesn't have

They are not the same.

5. $\mathcal{L}(b, w, \xi, \alpha, \beta) = \frac{1}{2}w^Tw + C\sum u_n \xi_n + \sum \alpha_n\left[\rho_n - \xi_n - y_n(w^T x_n + b)\right] + \sum \beta_n(-\xi_n)$

$\alpha_n, \beta_n \geq 0. \forall n.$

6. (i) $\frac{\partial L}{\partial b} = 0 = -\sum \alpha_n y_n$

$\Rightarrow \frac{1}{2}w^Tw + C\sum u_n\xi_n + \sum \alpha_n\left[\rho_n - \xi_n - y_n(w^T x_n)\right] + \beta_n(-\xi_n)$

(ii) $\frac{\partial L}{\partial w_i} = 0 = w_i - \sum \alpha_n y_n x_{n,i} \Rightarrow w = \sum \alpha_n y_n x_n$

$\Rightarrow \frac{1}{2}w^Tw + C\sum u_n\xi_n + \sum \alpha_n\left[\rho_n - \xi_n\right] - w^Tw + \sum \beta_n(-\xi_n)$

(iii) $\frac{\partial L}{\partial \xi_n} = 0 = -\alpha_n - \beta_n + C u_n \Leftrightarrow 0 \leq \alpha_n \leq C u_n.$

$\Rightarrow \frac{1}{2}w^Tw + \sum \alpha_n \rho_n - w^Tw$

Lagrange dual problem: $\max_{\substack{\alpha \\ (w=\sum \alpha_n y_n x_n) \\ 0\leq \alpha_n \leq C u_n \\ \sum y_n \alpha_n = 0}} \left(\min_{b, w, \xi} \frac{1}{2}w^Tw + \sum \alpha_n \rho_n - w^Tw \right)$

$$= \max_{\substack{\alpha \\ 0\leq \alpha_n \leq C u_n \\ \sum y_n \alpha_n = 0}} \left(-\frac{1}{2}\left\|\sum \alpha_n y_n x_n\right\|^2 + \sum \alpha_n \rho_n \right)$$

Scanned by CamScanner

7. (P_1')'s hard margin primal problem : $\max_{b',w'} \frac{1}{\|w'\|}$ in $\min y_n(w'^T x_n + b') = 0.25$

the equivalence ($\Leftrightarrow$) means they have the same optimal hyperplane

$w' = 0.25w$, $b' = 0.25b$ $\Leftrightarrow$ $\max_{b,w} \frac{0.25}{\|w\|}$ in $\min y_n(w'^T x_n + b') = 0.25$

$\Leftrightarrow$ $\max_{b,w} \frac{1}{\|w\|}$ in $\min y_n(w^T x_n + b) = 1$

This is (P_1)'s hard margin primal problem

Then we obtain (P_1')'s and (P)'s soft margin primal problem

if $\xi' = 0.25\xi$, (P_1')'s soft margin primal $\Leftrightarrow$ (P_1)'s soft margin primal.

$(b_*, w_*, \xi_*) = (4b_*', 4w_*', 4\xi_*')$

8. hard margin SVM dual : $\min_{\alpha} \frac{1}{2} \sum\sum \alpha_n \alpha_m y_n y_m z_n^T z_m - \sum \alpha_n$

subject to $\sum y_n \alpha_n = 0$, $0 \le \alpha_n$.

Soft margin SVM dual : $\min_{\alpha} \frac{1}{2} \sum\sum \alpha_n \alpha_m y_n y_m z_n^T z_m - \sum \alpha_n$

subject to $\sum y_n \alpha_n = 0$, $0 \le \alpha_n \le C$

The optimal solution in hard SVM also satisfies $0 \le \alpha_n^* \le C$ and $\sum y_n \alpha_n = 0$

if α^* is not the vector corresponded to the optimal solution of soft margin SVM, we have $\alpha^{*\prime}$ which satisfies $\sum y_n \alpha_n^{*\prime}$, $0 \le \alpha_n^{*\prime}$

such that $\frac{1}{2} \sum\sum \alpha_n^{*\prime} \alpha_m^{*\prime} y_n y_m z_n^T z_m - \sum \alpha_n^{*\prime} \le \frac{1}{2} \sum\sum \alpha_n^* \alpha_m^* y_n y_m z_n^T z_m - \sum \alpha_n^*$

$(\rightarrow\leftarrow)$

So, α^* must also be the vector corresponded to the optimal solu. of soft margin SVM.

Scanned by CamScanner

9. (b)(c)(d).

Solu. (a) let $K_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $x^T K_1 x \geq 0$

$$K = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix}$$

$$(-2\ 1)\begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix}\begin{pmatrix} -2 \\ 1 \end{pmatrix} = (0\ -2)\begin{pmatrix} -2 \\ 1 \end{pmatrix} = -2$$

So K is not positive semi-definite (p.s.d)

(b) $K = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \Rightarrow x^T A x = (\sum x_i)^2 \geq 0$

is p.s.d and symmetric

(c) $\frac{1}{2-k} = \frac{1}{2(1-\frac{k}{2})} = \frac{1}{2}(1 + \frac{k}{2} + (\frac{k}{2})^2 + \dots)$

claim: (i) $K+K'$ is p.s.d and symmetric, K and K' are valid kernel

(ii) $K \circ K'$ is p.s.d and symmetric, $\circ$ is hadamard product

pf. (i) $(K+K')^T = K^T + K'^T = K+K'$ (symmetric)

$$x^T(K+K')x = \underbrace{x^T K x}_{\geq 0} + \underbrace{x^T K' x}_{\geq 0} \geq 0 \quad (\text{p.s.d})$$

(ii) $(K \circ K')^T = K^T \circ K'^T = K \circ K'$ (symmetric)

By Schur product Thm $K \circ K'$ is p.s.d.

pf. A, B p.s.d $\Rightarrow A = \sum \mu_i a_i a_i^T$, $B = \sum v_i b_i b_i^T$

$\{\mu_i\}_{i=1..N}$ are A's eigenvalues

$\{v_i\}_{i=1..N}$ are B's eigenvalues

$$A \circ B = \sum_{ij} \mu_i v_j (a_i a_i^T) \circ (b_j b_j^T) \quad \mu_i, v_j \geq 0, \forall i, j$$

$$= \sum_{ij} \mu_i v_j (a_i \circ b_j)(a_i \circ b_j)$$

$$x^T A \circ B x = x^T\left(\sum_{ij}(a_i \circ b_j)(a_i \circ b_j)^T\right)x$$

$$= \sum_{ij}\left(x^T (a_i \circ b_j)(a_i \circ b_j) x\right)$$

k-th element of a_i k-th of b_j

$x^T(a_i \circ b_j)(a_i \circ b_j)x = (\sum_k m_{ik} n_{jk} x_k^2)^2 \geq 0$, So $x^T A \circ B x \geq 0$

16. Prove!

Original: $\min_{\alpha} \sum\sum \alpha_n \alpha_m y_n y_m K(x_n, x_m) - \sum \alpha_n$ subject to $\sum y_n \alpha_n = 0$, $0 \le \alpha_n \le C$

$$\Leftrightarrow \min_{\alpha} \sum\sum \frac{1}{p} \alpha_n \alpha_m y_n y_m K(x_n, x_m) - \frac{1}{p} \sum \alpha_n, \quad \sum y_n \alpha_n = 0,\ 0 \le \alpha_n \le C$$

let $\beta_n = \frac{\alpha_n}{p}$
$$\Leftrightarrow \min_{p\beta_n} \sum\sum \frac{1}{p} p\beta_n p\beta_m y_n y_m K(x_n, x_m) - \frac{1}{p} \sum p\beta_n, \quad \sum y_n p\beta_n = 0,\ 0 \le p\beta_n \le C$$

$$\Leftrightarrow \min_{\beta_n} \sum\sum p \beta_n \beta_m y_n y_m K(x_n, x_m) - \sum \beta_n, \quad \sum y_n \beta_n = 0,\ 0 \le \beta_n \le \frac{C}{p}$$

This is dual of soft-margin SVM with $\tilde{K} = pK(x, x')$, $\tilde{C} = \frac{C}{p}$

17. hard margin SVM: $\max_{b,w} \left(\min_{y_n, x_n} \frac{1}{\|w\|} y_n (w^T x_n + b) \right)$ subject to $y_n (w^T x_n + b) > 0$

$y_n (w^T x_n + b)$ can be the same no matter what w_0 is ($\because$ b can change).

if $y_n (w^T x_n + b)$ is fixed, we then want to choose (b, w) for maximization

To do this, we minimize $\|w\|$ by letting $w_0 = 0$

Then we derive soft margin SVM by hard margin SVM, w_0 is also zero.

18.

$$\min \frac{1}{N} \sum (y_n - w^T x_n)^2 \quad \text{s.t. } w^T w \le C$$

$$\Rightarrow \frac{1}{N} (y - Xw)^2 + \lambda (w^T w - C)$$

$$\Rightarrow \frac{1}{N} (y^T - w^T X^T)(y - Xw) + \lambda (w^T w - C)$$

$$\Rightarrow \frac{1}{N} (y^T y - 2 w^T X^T y + w^T X^T X w) + \lambda (w^T w - C) \quad \text{denoted as } f$$

$$0 = \frac{\partial f}{\partial w} = \frac{1}{N} (-2X^T y + 2X^T X w) + 2\lambda w$$

$$\Leftrightarrow X^T y = (X^T X + N\lambda I) w.$$