CSIE 5432/5433 — Machine Learning Foundations/Techniques

Name: 李吉昌 **Homework 5**

Student Number: r08922a27 **Due Date:** December 25 2020, 13:00

Hard-Margin SVM and Large Margin

1. Answer: [d]

令 $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$, $\mathbf{w}^*$ 為一解, 則必滿足 $\begin{cases} w_1 - 2w_2 + 4w_3 + b \leq -1 & (1) \\ w_1 + b \geq 1 & (2) \\ w_1 + 2w_2 + 4w_3 + b \leq -1 & (3) \end{cases}$

由 (1)(2) 和 (1)(3) 可得 $\begin{cases} +2w_2 - 4w_3 \geq 1 & (4) \\ -2w_2 - 4w_3 \geq 1 & (5) \end{cases}$, 由 (4)(5) 可推至 $w_3 \leq -\frac{1}{2}$ 和 $w_2 = 0$,

代回 (1) 可得 $w_1 + b \leq 1$, 綜上所敘, 需同時滿足 $w_1 + b \leq 1$ 以及 $w_1 + b \geq 1$, 則 $w_1 + b = 1$。

$\frac{1}{2}\mathbf{w}^T\mathbf{w} = \frac{1}{2}(w_1^2 + w_2^2 + w_3^2) \geq \frac{1}{2}(w_2^2 + w_3^2)$ 恆成立, 得最佳解 $\mathbf{w}^*$ 的 $w_1^* = 0$ 且 $b^* = 1$。

2. Answer: [b]

由上題已知, $w_2^* = 0$, $w_1^* = 0$, $b^* = 1$, 則 $\frac{1}{2}\mathbf{w}^T\mathbf{w} = \frac{1}{2}w_3^2$。

因為 $w_3 \leq -\frac{1}{2}$, 得 $\frac{1}{2}\mathbf{w}^T\mathbf{w} \leq \frac{1}{8}$, $(\mathbf{w}^*)^T\mathbf{w}^* = \frac{1}{4}$, $\text{margin}(\mathbf{b}^*, \mathbf{w}^*) = \frac{1}{\|\mathbf{w}^*\|} = \frac{1}{\frac{1}{2}} = 2$。

3. Answer: [e]

令 $h(x) = \text{sign}(x - \theta)$, 若 h 為一 feasible solution 表示 h 能使 $E_{in}(h) = 0$, 如果 θ 不在 x_M 和 x_{M+1} 之間, 必存在至少一 $y_n = -1$ 且 $x_n > \theta$ 或 $y_n = +1$ 且 $x_n < \theta$ 的情況使得 $E_{in}(h) \neq 0$, 因此 θ 必介於 x_M 和 x_{M+1}, 得 $\text{margin}(\theta) = \min_i |x_i - \theta| = \min(|x_M - \theta|, |x_{M+1} - \theta|)$。

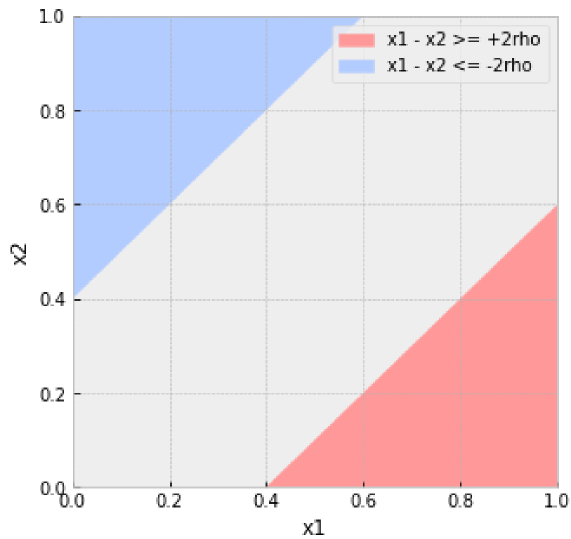
令一常數 $\epsilon \in [-\frac{1}{2}(x_{M+1} - x_M), \frac{1}{2}(x_{M+1} - x_M))$, 使得 $\theta = \frac{1}{2}(x_{M+1} + x_M) + \epsilon \in [x_M, x_{M+1})$。
當 $\epsilon \geq 0$ 時, θ 較接近 x_{M+1}, 則 $\text{margin}(\theta) = \frac{1}{2}(x_{M+1} - x_M) - \epsilon$; 當 $\epsilon < 0$, θ 較接近 x_M, 則 $\text{margin}(\theta) = \frac{1}{2}(x_{M+1} - x_M) + \epsilon$。
綜上所述, $\text{margin}(\theta) = \frac{1}{2}(x_{M+1} - x_M) - |\theta| \leq \frac{1}{2}(x_{M+1} - x_M)$, 當 $\epsilon = 0$ 時, $\theta = \frac{1}{2}(x_M + x_{M+1})$ 為最佳解, 得最大值 $\text{margin}(\theta) = \frac{1}{2}(x_{M+1} - x_M)$

4. Answer: [a]

dichotomy 所有可能為 $\mathcal{H} = \{⊚⊚, ××, ⊚×, ×⊚\}$, 分別對應的數量為 1, 根據題意, 令指定的 dichotomy 為 h,「expected number of dichotomies」可寫作 $\sum_{h \in \mathcal{H}} 1 \cdot P(h)$, $P(h)$ 可分成 $\{⊚⊚, ××\}$ 和 $\{⊚×, ×⊚\}$ 兩種情況討論:

當 $h \in \{⊚⊚, ××\}$ 時, 因為 threshold 沒有區間限制, threshold 只要設大於 $1 + \rho$/小於 $0 - \rho$ 必然會發生 ××/⊚⊚ 被 shatter 的情況, 得 $P(h) = P[\![⊚⊚]\!] = P[\![××]\!] = 1$。

當 $h \in \{⊚×, ×⊚\}$ 時, 只要在 $|x_1 - x_2| \geq 2\rho$ 的情況, 必可找到 threshold 將 x_1 和 x_2 劃分為不同類, 得 $P(h) = P[\![⊚×]\!] = P[\![×⊚]\!] = P[\![|x_1 - x_2| \geq 2\rho]\!]$, 令 $x1$ 和 $x2$ 的機率密度為 $f_{X_1,X_2}(x1, x2)$, 密度積分面積如下圖所示(以 $\rho = 0.2$ 為例):



$P[\![|x_1 - x_2| \geq 2\rho]\!] = \iint\limits_{|x_1-x_2|\geq 2\rho} f_{X_1,X_2}(x1, x2)dx_1dx_2$, x_1 和 x_2 互相獨立且為介於 $[0, 1]$ 的平均分布,

$$= \iint\limits_{|x_1-x_2|\geq 2\rho} (\frac{1}{1-0} \cdot \frac{1}{1-0})dx_1dx_2 = \frac{1}{2}(1-2\rho)^2 + \frac{1}{2}(1-2\rho)^2 = (1-2\rho)^2$$

綜上所述, $\sum_{h\in\mathcal{H}} 1 \cdot P(h) = P[\![\circledcirc\circledcirc]\!] + P[\![\times\times]\!] + P[\![\circledcirc\times]\!] + P[\![\times\circledcirc]\!] = 2 + 2(1-2\rho)^2$

Dual Problem of Quadratic Programming

5. Answer: [c]

根據題意, uneven-margin SVM 目標 $\max\limits_{\text{all } \alpha_n\geq 0} \min\limits_{b,\mathbf{w}} \mathcal{L}(b, \mathbf{w}, \boldsymbol{\alpha})$ 可以寫作:

$$\max_{\text{all } \alpha_n\geq 0} \min_{b,\mathbf{w}} \{\tfrac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{n=1}^{N} \alpha_n([\![y_n = +1]\!](\rho_+ - y_n(\mathbf{w}^T\mathbf{x}_n + b)) + [\![y_n = -1]\!](\rho_- - y_n(\mathbf{w}^T\mathbf{x}_n + b)))\}$$

當 $\frac{\partial\mathcal{L}(b,\mathbf{w},\boldsymbol{\alpha})}{\partial b} = 0$ 時:

$$\sum_{n=1}^{N} \alpha_n([\![y_n = +1]\!]y_n + [\![y_n = -1]\!]y_n) = \sum_{n=1}^{N} \alpha_n y_n = 0$$

當 $\frac{\partial\mathcal{L}(b,\mathbf{w},\boldsymbol{\alpha})}{\partial \mathbf{w}} = 0$ 時:

$$\mathbf{w} + \sum_{n=1}^{N} \alpha_n([\![y_n = +1]\!](-y_n\mathbf{x}_n) + [\![y_n = -1]\!](-y_n\mathbf{x}_n)) = \mathbf{w} - \sum_{n=1}^{N} \alpha_n y_n \mathbf{x}_n = 0, \mathbf{w} = \sum_{n=1}^{N} \alpha_n y_n \mathbf{x}_n$$

將 $\sum_{n=1}^{N} \alpha_n y_n = 0$ 和 $\mathbf{w} = \sum_{n=1}^{N} \alpha_n y_n \mathbf{x}_n$ 兩限制代入 $\max\limits_{\text{all } \alpha_n\geq 0} \min\limits_{b,\mathbf{w}} \mathcal{L}(b, \mathbf{w}, \boldsymbol{\alpha})$:

$$
\begin{aligned}
\max_{\text{all } \alpha_n \geq 0} \min_{b,\mathbf{w}} \mathcal{L}(b, \mathbf{w}, \boldsymbol{\alpha}) = \max_{\text{all } \alpha_n \geq 0} \{ & \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m \mathbf{x}_n^T \mathbf{x}_m \\
& + \sum_{n=1}^{N} \alpha_n ([\![y_n = +1]\!](\rho_+ - \sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m \mathbf{x}_n^T \mathbf{x}_m) \\
& + [\![y_n = -1]\!](\rho_- - \sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m \mathbf{x}_n^T \mathbf{x}_m))\} \\
= \max_{\text{all } \alpha_n \geq 0} \{ & -\frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m \mathbf{x}_n^T \mathbf{x}_m \\
& + \sum_{n=1}^{N} \alpha_n [\![y_n = +1]\!]\rho_+ + \sum_{n=1}^{N} \alpha_n [\![y_n = -1]\!]\rho_- \} \\
= \min_{\text{all } \alpha_n \geq 0} \{ & \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m \mathbf{x}_n^T \mathbf{x}_m \\
& - \sum_{n=1}^{N} \alpha_n [\![y_n = +1]\!]\rho_+ - \sum_{n=1}^{N} \alpha_n [\![y_n = -1]\!]\rho_- \}
\end{aligned}
$$

6. Answer: [e]

令 $\boldsymbol{\alpha}, \mathbf{w}, b$ 為 $\rho_+ = \rho_- = 1$ 時的 feasible solution, $\boldsymbol{\alpha}' = \frac{\rho_+ + \rho_-}{2} \boldsymbol{\alpha}$,
則 $\mathbf{w}' = \sum_{n=1}^{N} \alpha_n' y_n \mathbf{x}_n = \frac{\rho_+ + \rho_-}{2} \sum_{n=1}^{N} \alpha_n y_n \mathbf{x}_n = \frac{\rho_+ + \rho_-}{2} \mathbf{w}$。

因為 $\frac{\rho_+ + \rho_-}{2} \geq 0$ 和 $\alpha_n \geq 0$, 因此 $\alpha_n' \geq 0$ 亦成立, 符合 dual feasible 的條件。

因為 $\sum_{n=1}^{N} \alpha_n y_n = 0$, $\sum_{n=1}^{N} \alpha_n' y_n = \frac{\rho_+ + \rho_-}{2} \sum_{n=1}^{N} \alpha_n y_n = \frac{\rho_+ + \rho_-}{2} \cdot 0 = 0$, 符合 dual-inner optimal 的條件。

為符合 primal feasible 以及 complementary slackness, 令 $b' = \frac{\rho_+ + \rho_-}{2} b + \frac{\rho_+ - \rho_-}{2}$, 共可依據 $y_n = +1/-1$ 和 $y_n(\mathbf{w}^T \mathbf{x}_n + b) > 1 / y_n(\mathbf{w}^T \mathbf{x}_n + b) = 1$ 分成四個狀況討論:

當 $y_n = +1$ 且 $y_n(\mathbf{w}^T \mathbf{x}_n + b) = 1$ 時:
$y_n(\mathbf{w}'^T \mathbf{x}_n + b') = \frac{\rho_+ + \rho_-}{2} y_n(\mathbf{w}^T \mathbf{x}_n + b) + \frac{\rho_+ - \rho_-}{2} y_n = \frac{\rho_+ + \rho_-}{2} \cdot 1 + \frac{\rho_+ - \rho_-}{2} \cdot 1 = \rho_+$

當 $y_n = +1$ 且 $y_n(\mathbf{w}^T \mathbf{x}_n + b) > 1$ 時:
$y_n(\mathbf{w}'^T \mathbf{x}_n + b') = \frac{\rho_+ + \rho_-}{2} y_n(\mathbf{w}^T \mathbf{x}_n + b) + \frac{\rho_+ - \rho_-}{2} y_n > \frac{\rho_+ + \rho_-}{2} \cdot 1 + \frac{\rho_+ - \rho_-}{2} \cdot 1 = \rho_+$

當 $y_n = -1$ 且 $y_n(\mathbf{w}^T \mathbf{x}_n + b) = 1$ 時:
$y_n(\mathbf{w}'^T \mathbf{x}_n + b') = \frac{\rho_+ + \rho_-}{2} y_n(\mathbf{w}^T \mathbf{x}_n + b) + \frac{\rho_+ - \rho_-}{2} y_n = \frac{\rho_+ + \rho_-}{2} \cdot 1 + \frac{\rho_+ - \rho_-}{2} \cdot -1 = \rho_-$

當 $y_n = -1$ 且 $y_n(\mathbf{w}^T \mathbf{x}_n + b) > 1$ 時:
$y_n(\mathbf{w}'^T \mathbf{x}_n + b') = \frac{\rho_+ + \rho_-}{2} y_n(\mathbf{w}^T \mathbf{x}_n + b) + \frac{\rho_+ - \rho_-}{2} y_n > \frac{\rho_+ + \rho_-}{2} \cdot 1 + \frac{\rho_+ - \rho_-}{2} \cdot -1 = \rho_-$

綜上所述, 當 $\alpha' = \frac{\rho_+ + \rho_-}{2} \alpha$, $\mathbf{w}' = \frac{\rho_+ + \rho_-}{2} \mathbf{w}$ 且令 $b' = \frac{\rho_+ + \rho_-}{2} b + \frac{\rho_+ - \rho_-}{2}$ 時, 符合 uneven-margin SVM 的 KKT 條件, 必為 feasible solution。

令 α^* 為 α 的 optimal solution, 則:

$$\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha'_n\alpha'_m y_n y_m = (\frac{\rho_+ + \rho_-}{2})^2\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m \geq (\frac{\rho_+ + \rho_-}{2})^2\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha^*_n\alpha^*_m y_n y_m$$
$$= \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}(\frac{\rho_+ + \rho_-}{2}\alpha^*_n)(\frac{\rho_+ + \rho_-}{2}\alpha^*_m)y_n y_m$$

$\frac{\rho_+ + \rho_-}{2}\alpha^*$ 恆為可以更接近目標的 feasible solution, 得 α' 的 optimal solution 為 $\frac{\rho_+ + \rho_-}{2}\alpha^*$。

7. Answer: [d]

令 $K(\mathbf{x}, \mathbf{x}')$ 對應到的 Gram matrix 為 $\begin{pmatrix}\frac{1}{2}, \frac{1}{2}\\ \frac{1}{2}, \frac{1}{2}\end{pmatrix}$, det $\begin{pmatrix}\frac{1}{2}-\lambda, \frac{1}{2}\\ \frac{1}{2}, \frac{1}{2}-\lambda\end{pmatrix} = 0$, $\lambda(\lambda-1) = 0$, eigenvalue 為 0, 1。
則 $\log_2 K(\mathbf{x}, \mathbf{x}')$ 對應到的 Gram matrix 為 $\begin{pmatrix}-1, -1\\ -1, -1\end{pmatrix}$, det $\begin{pmatrix}-1-\lambda, -1\\ -1, -1-\lambda\end{pmatrix} = 0$, $\lambda(\lambda+2) = 0$, eigenvalue 為 0, -2。

因為 $\log_2 K(\mathbf{x}, \mathbf{x}')$ 的 Gram matrix 的 eigenvalue 存在小於 0 的可能, 不符合 positive semi-definite 的性質, 因此 $\log_2 K(\mathbf{x}, \mathbf{x}')$ 不具備 valid kernel 的條件。

8. Answer: [c]

$\|\boldsymbol{\phi}(\mathbf{x}) - \boldsymbol{\phi}(\mathbf{x}')\|^2 = \boldsymbol{\phi}(\mathbf{x})^T\boldsymbol{\phi}(\mathbf{x}) - \boldsymbol{\phi}(\mathbf{x})^T\boldsymbol{\phi}(\mathbf{x})' - \boldsymbol{\phi}(\mathbf{x})'^T\boldsymbol{\phi}(\mathbf{x}) + \boldsymbol{\phi}(\mathbf{x})'^T\boldsymbol{\phi}(\mathbf{x})'$

當 $\mathbf{x} - \mathbf{x}' = 0$, $\exp(\gamma\|\mathbf{x} - \mathbf{x}'\|^2) = 1$, 為極大值; 當 $\mathbf{x} - \mathbf{x}' \to \infty$, $\exp(\gamma\|\mathbf{x} - \mathbf{x}'\|^2) \to 0$, 為極小值,
根據題意代入 $\boldsymbol{\phi}(\mathbf{x}_1)^T\boldsymbol{\phi}(\mathbf{x}_2) = \exp(\gamma\|\mathbf{x}_1 - \mathbf{x}_2\|^2)$, 可得 $\|\boldsymbol{\phi}(\mathbf{x}) - \boldsymbol{\phi}(\mathbf{x}')\|^2 = 2 - 2\exp(\gamma\|\mathbf{x} - \mathbf{x}'\|^2) \leq 2$。
由上述可知, $\|\boldsymbol{\phi}(\mathbf{x}) - \boldsymbol{\phi}(\mathbf{x}')\|^2$ 有可能發生超過 1 的情況, 因此 2 為「the tightest upper bound」。

9. Answer: [d]

根據題意, 令 $h_{\boldsymbol{\alpha},b}(\mathbf{x}_n) = \text{sign}(\sum_{m=1}^{N} y_m K(\mathbf{x}_m, \mathbf{x}_n))$, 若 $E_{in}(h) = 0$, 表示所有 $h_{\boldsymbol{\alpha},b}(\mathbf{x}_n) = y_n$ 皆成立。

(1) 當 $y_n = +1$ 時:

$$\sum_{m=1}^{N} y_m K(\mathbf{x}_m, \mathbf{x}_n) = \sum_{\substack{m=1\\ m\neq n}}^{N} y_m K(\mathbf{x}_m, \mathbf{x}_n) + 1 \geq \sum_{\substack{m=1\\ m\neq n}}^{N} (-1)\cdot K(\mathbf{x}_m, \mathbf{x}_n) + 1$$
$$\geq \sum_{\substack{m=1\\ m\neq n}}^{N} (-1)\cdot \exp(-\gamma\epsilon^2) + 1 = -(N-1)\exp(-\gamma\epsilon^2) + 1$$

$\sum_{m=1}^{N} y_m K(\mathbf{x}_m, \mathbf{x}_n) > 0$ 皆成立代表 $y_n = +1$ 的情況都會答對, 若 $-(N-1)\exp(-\gamma\epsilon^2) + 1 > 0$ 成立, 則 $\sum_{m=1}^{N} y_m K(\mathbf{x}_m, \mathbf{x}_n) > 0$ 必成立, 整理移項後, 得 $(N-1)\exp(-\gamma\epsilon^2) < 1$。

(2) 當 $y_n = -1$ 時:

$$\sum_{m=1}^{N} y_m K(\mathbf{x}_m, \mathbf{x}_n) = \sum_{\substack{m=1 \\ m\neq n}}^{N} y_m K(\mathbf{x}_m, \mathbf{x}_n) - 1 \leq \sum_{\substack{m=1 \\ m\neq n}}^{N} (+1)\cdot K(\mathbf{x}_m, \mathbf{x}_n) - 1$$

$$\leq \sum_{\substack{m=1 \\ m\neq n}}^{N} (+1)\cdot \exp(-\gamma\epsilon^2) - 1 = (N-1)\exp(-\gamma\epsilon^2) - 1$$

$\sum\limits_{m=1}^{N} y_m K(\mathbf{x}_m, \mathbf{x}_n) \leq 0$ 皆成立代表 $y_n = -1$ 的情況都會答對, 若 $(N-1)\exp(-\gamma\epsilon^2) - 1 \leq 0$ 成立, 則 $\sum\limits_{m=1}^{N} y_m K(\mathbf{x}_m, \mathbf{x}_n) \leq 0$ 必成立, 整理移項後, 得 $(N-1)\exp(-\gamma\epsilon^2) \leq 1$。

由 (1)(2) 得, $(N-1)\exp(-\gamma\epsilon^2) \leq 1$ 成立時, $E_{in}(h) = 0$ 必成立, 將其取對數並移項整理, 得 $\gamma \geq \frac{\ln(N-1)}{\epsilon^2}$, 且 $\frac{\ln(N-1)}{\epsilon^2}$ 為最小選項, 因此 $\frac{\ln(N-1)}{\epsilon^2}$ 為「the tightest lower bound」。

Kernel Perceptron Learning Algorithm

10. Answer: [c]

由題意可知 $\mathbf{w}_{t+1} = \mathbf{w}_t + [\![y_{n(t)} \neq \mathrm{sign}(\mathbf{w}_t^T \boldsymbol{\phi}(\mathbf{x}_{n(t)}))]\!] \cdot y_{n(t)} \boldsymbol{\phi}(\mathbf{x}_{n(t)})$, 代入 $\mathbf{w}_t = \sum\limits_{n=1}^{N} \alpha_{t,n(t)} \boldsymbol{\phi}(\mathbf{x}_{n(t)})$,

得 $\mathbf{w}_{t+1} = \sum\limits_{n=1}^{N} (\alpha_{t,n(t)} + [\![y_{n(t)} \neq \mathrm{sign}(\mathbf{w}_t^T \boldsymbol{\phi}(\mathbf{x}_{n(t)}))]\!] \cdot y_{n(t)}) \boldsymbol{\phi}(\mathbf{x}_{n(t)}) = \sum\limits_{n=1}^{N} \alpha_{t+1,n(t)} \boldsymbol{\phi}(\mathbf{x}_{n(t)})$,

綜上所述, 參數更新式可寫作 $\boldsymbol{\alpha}_{t+1} \leftarrow \boldsymbol{\alpha}_t$ except $\alpha_{t+1,n(t)} \leftarrow \alpha_{t,n(t)} + y_{n(t)}$。

11. Answer: [a]

根據題意代入 $\mathbf{w}_t = \sum\limits_{n=1}^{N} \alpha_{t,n} \boldsymbol{\phi}(\mathbf{x}_n)$, $\mathbf{w}_t^T \boldsymbol{\phi}(\mathbf{x}) = \sum\limits_{n=1}^{N} \alpha_{t,n} \boldsymbol{\phi}(\mathbf{x}_n)^T \boldsymbol{\phi}(\mathbf{x}) = \sum\limits_{n=1}^{N} \alpha_{t,n} K(\mathbf{x}_n, \mathbf{x})$

12. Answer: [b]

根據題意, 因為所有 $\mathbf{x}_n$ 皆為 bounded SV, 對所有 $\mathbf{x}_n$ 需滿足兩條件: $\begin{cases} y_n - y_n\xi_n = \mathbf{w}^T\mathbf{x}_n + b \\ \xi_n \geq 0 \end{cases}$

(1) 當 $y_n = +1$ 時:

$y_n - y_n\xi_n = 1 - \xi_n = \mathbf{w}^T\mathbf{x}_n + b$, 整理移項, 得 $\xi_n = 1 - (\mathbf{w}^T\mathbf{x}_n + b)$, 因為所有 $y_n = +1$ 的 $\mathbf{x}_n$ 滿足 $\xi_n \geq 0$ 的條件等價於 $\min\limits_{n:y_n>0} \xi_n \geq 0$, 將其代入 $\xi_n = 1 - (\mathbf{w}^T\mathbf{x}_n + b)$, 則 $\min\limits_{n:y_n>0} \{1 - (\mathbf{w}^T\mathbf{x}_n + b)\} \geq 0$,

可得 $b \leq \min\limits_{n:y_n>0} \{1 - \mathbf{w}^T\mathbf{x}_n\} = \min\limits_{n:y_n>0} \{1 - \sum\limits_{m=1}^{N} y_m\alpha_m K(\mathbf{x}_m, \mathbf{x}_n)\}$

(2) 當 $y_n = -1$ 時:

$y_n - y_n\xi_n = \xi_n - 1 = \mathbf{w}^T\mathbf{x}_n + b$, 整理移項, 得 $\xi_n = 1 + \mathbf{w}^T\mathbf{x}_n + b$, 因為所有 $y_n = -1$ 的 $\mathbf{x}_n$ 滿足 $\xi_n \geq 0$ 的條件等價於 $\min\limits_{n:y_n<0} \xi_n \geq 0$, 將其代入 $\xi_n = 1 + \mathbf{w}^T\mathbf{x}_n + b$, 則 $\min\limits_{n:y_n<0} \{1 + \mathbf{w}^T\mathbf{x}_n + b\} \geq 0$,

可得 $b \geq -\min\limits_{n:y_n<0} \{1 + \mathbf{w}^T\mathbf{x}_n\} = -\min\limits_{n:y_n<0} \{1 + \sum\limits_{m=1}^{N} y_m\alpha_m K(\mathbf{x}_m, \mathbf{x}_n)\}$

由 (1)(2) 可推得 b 的合法範圍, $-\min_{n:y_n<0}\{1+\sum_{m=1}^{N} y_m\alpha_m K(\mathbf{x}_m,\mathbf{x}_n)\} \le b \le \min_{n:y_n>0}\{1-\sum_{m=1}^{N} y_m\alpha_m K(\mathbf{x}_m,\mathbf{x}_n)\}$, b 最大值為 $\min_{n:y_n>0}\{1-\sum_{m=1}^{N} y_m\alpha_m K(\mathbf{x}_m,\mathbf{x}_n)\}$。

13. Answer: [e]

根據題意, SVM 目標 $\max_{\text{all }\alpha_n\ge 0}\min_{b,\mathbf{w},\boldsymbol{\xi}} \mathcal{L}(b,\mathbf{w},\boldsymbol{\xi},\boldsymbol{\alpha})$ 可以寫作:

$$\max_{\text{all }\alpha_n\ge 0}\min_{b,\mathbf{w},\boldsymbol{\xi}}\{\tfrac{1}{2}\mathbf{w}^T\mathbf{w}+C\cdot\sum_{n=1}^{N}\xi_n^2+\sum_{n=1}^{N}\alpha_n(1-\xi_n-y_n(\mathbf{w}^T\mathbf{x}_n+b))\}$$

當 $\frac{\partial\mathcal{L}(b,\mathbf{w},\boldsymbol{\xi},\boldsymbol{\alpha})}{\partial b}=0$ 時, $\sum_{n=1}^{N}\alpha_n y_n=0$; 當 $\frac{\partial\mathcal{L}(b,\mathbf{w},\boldsymbol{\xi},\boldsymbol{\alpha})}{\partial\mathbf{w}}=0$ 時, $\mathbf{w}-\sum_{n=1}^{N}\alpha_n y_n\mathbf{x}_n=0, \mathbf{w}=\sum_{n=1}^{N}\alpha_n y_n\mathbf{x}_n$

當 $\frac{\partial\mathcal{L}(b,\mathbf{w},\boldsymbol{\xi},\boldsymbol{\alpha})}{\partial\xi_n}=0$ 時, $2C\cdot\xi_n-\alpha_n=0$, $\xi_n=\frac{\alpha_n}{2C}$。

將上述三個條件代回 $\max_{\text{all }\alpha_n\ge 0}\min_{b,\mathbf{w},\boldsymbol{\xi}} \mathcal{L}(b,\mathbf{w},\boldsymbol{\xi},\boldsymbol{\alpha})$ 可得:

$$\begin{aligned}
\max_{\text{all }\alpha_n\ge 0}\min_{b,\mathbf{w},\boldsymbol{\xi}} \mathcal{L}(b,\mathbf{w},\boldsymbol{\xi},\boldsymbol{\alpha}) &= \max_{\text{all }\alpha_n\ge 0}\{-(\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m K(\mathbf{x}_n,\mathbf{x}_m)+C\cdot\sum_{n=1}^{N}\frac{\alpha_n^2}{4C^2}-\sum_{n=1}^{N}\alpha_n)\} \\
&= \min_{\text{all }\alpha_n\ge 0}\{\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m K(\mathbf{x}_n,\mathbf{x}_m)+C\cdot\sum_{n=1}^{N}\frac{\alpha_n^2}{4C^2}-\sum_{n=1}^{N}\alpha_n\} \\
&= \min_{\text{all }\alpha_n\ge 0}\{\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m K(\mathbf{x}_n,\mathbf{x}_m)+\frac{1}{2}\sum_{n=1}^{N}\frac{\alpha_n^2}{2C}-\sum_{n=1}^{N}\alpha_n\} \\
&= \min_{\text{all }\alpha_n\ge 0}\{\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m K(\mathbf{x}_n,\mathbf{x}_m) \\
&\quad +\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m\frac{1}{2C}[\![n=m]\!]-\sum_{n=1}^{N}\alpha_n\} \\
&= \min_{\text{all }\alpha_n\ge 0}\{\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m(K(\mathbf{x}_n,\mathbf{x}_m)+\frac{1}{2C}[\![n=m]\!])-\sum_{n=1}^{N}\alpha_n\}
\end{aligned}$$

14. Answer: [c]

由上題可知, 為滿足 $\min_{b,\mathbf{w},\boldsymbol{\xi}} \mathcal{L}(b,\mathbf{w},\boldsymbol{\xi},\boldsymbol{\alpha})$ 的條件, solution 必滿足 $\xi_n=\frac{\alpha_n}{2C}$, 因此 $\boldsymbol{\xi}^*=\frac{1}{2C}\boldsymbol{\alpha}^*$ 恆成立。

Experiment

程式碼實作細節如下, 可以透過 parser 的 −−tra_path/−−tst_path 設置訓練資料和測試資料的路徑

`python code.py --tra_path satimage.scale --tst_path satimage.scale.t`

```python
from svmutil import *
import numpy as np
import argparse

'''Define Function'''
```

```python
def label_filtering(Y, label):
    Y_tmp = np.zeros(Y.shape)
    Y_tmp[Y == label] = 1
    Y_tmp[Y != label] = -1
    return Y_tmp


def feature_processing(X):
    X_tmp = []
    for x in X:
        x_tmp = np.zeros(36)
        for i in x:
            x_tmp[i - 1] = x[i]
        X_tmp.append(x_tmp)
    return np.array(X_tmp)


def get_data(path):
    Y, X = svm_read_problem(path)
    return np.array(Y), feature_processing(X)


def main():
    '''Parsing'''
    parser = argparse.ArgumentParser(
        description='Argument Parser for ML HW5.')

    parser.add_argument('--tra_path', default='satimage.scale')
    parser.add_argument('--tst_path', default='satimage.scale.t')
    args = parser.parse_args()

    # load data
    Y_tra_mul, X_tra = get_data(args.tra_path)
    Y_tst_mul, X_tst = get_data(args.tst_path)

    '''Answer questions'''
    print('RUNNING Q15...')
    Y_tra = label_filtering(Y_tra_mul, 3)
    model = svm_train(Y_tra, X_tra, '-s 0 -t 0 -c 10 -q')
    alpha = np.array([a[0] for a in model.get_sv_coef()]).reshape(-1, 1)
    SV = feature_processing(model.get_SV())
    w = np.dot(SV.T, alpha)
    print('Answer of Q15 : {:.4f}\n'.format(np.linalg.norm(w)))

    print('RUNNING Q16...')
    SV_num_list = []
    min_Ein = 1.0
    min_cls = 1
    for l in [1, 2, 3, 4, 5]:
```

```
    Y_tra = label_filtering(Y_tra_mul, l)
    model = svm_train(Y_tra, X_tra, '-s 0 -t 1 -g 1 -r 1 -d 2 -c 10 -q')
    SV_num_list.append(model.get_nr_sv())
    _, pre_acc, _ = svm_predict(Y_tra, X_tra, model, '-q')
    Ein = 1 - 0.01 * pre_acc[0]
    if Ein < min_Ein:
        min_Ein = Ein
        min_cls = l
print('Answer of Q16 : \{:d}" versus \not {:d}"\n'.format(
    min_cls, min_cls))

print('RUNNING Q17...')
print('Answer of Q17 : {:d}\n'.format(max(SV_num_list)))

print('RUNNING Q18...')
Y_tra = label_filtering(Y_tra_mul, 6)
Y_tst = label_filtering(Y_tst_mul, 6)
gamma = 10
min_Eout = 1.0
min_C = 1e-2
for C in [1e-2, 1e-1, 1e0, 1e1, 1e2]:
    model = svm_train(
        Y_tra, X_tra, '-s 0 -t 2 -g {:f} -c {:f} -q'.format(gamma, C))
    _, pre_acc, _ = svm_predict(Y_tst, X_tst, model, '-q')
    Eout = 1 - 0.01 * pre_acc[0]
    if Eout < min_Eout:
        min_Eout = Eout
        min_C = C
print('Answer of Q18 : {:.2f}\n'.format(min_C))

print('RUNNING Q19...')
C = 0.1
min_Eout = 1.0
min_gamma = 1e-1
for gamma in [1e-1, 1e0, 1e1, 1e2, 1e3]:
    model = svm_train(
        Y_tra, X_tra, '-s 0 -t 2 -g {:f} -c {:f} -q'.format(gamma, C))
    _, pre_acc, _ = svm_predict(Y_tst, X_tst, model, '-q')
    Eout = 1 - 0.01 * pre_acc[0]
    if Eout < min_Eout:
        min_Eout = Eout
        min_gamma = gamma
print('Answer of Q19 : {:.2f}\n'.format(min_gamma))

print('RUNNING Q20...')
gamma_list = [1e-1, 1e0, 1e1, 1e2, 1e3]
best_counter = np.zeros(len(gamma_list))
C = 0.1
for _ in range(1000):
    randomlist = np.random.permutation(len(X_tra))
    min_Eout = 1.0
```

```python
        min_gamma_idx = 0
        for i in range(len(gamma_list)):
            gamma = gamma_list[i]
            model = svm_train(Y_tra[randomlist[200:]], X_tra[randomlist[200:]],
                              '-s 0 -t 2 -g {:f} -c {:f} -q'.format(gamma, C))
            _, pre_acc, _ = svm_predict(
                Y_tra[randomlist[:200]], X_tra[randomlist[:200]], model, '-q')
            Eout = 1 - 0.01 * pre_acc[0]
            if Eout < min_Eout:
                min_Eout = Eout
                min_gamma_idx = i
        best_counter[min_gamma_idx] += 1
    print('Answer of Q20 : {:.2f}\n'.format(
        gamma_list[np.argmax(best_counter)]))


if __name__ == "__main__":
    main()
```

15. Ans: [d]	16. Ans: [b]	17. Ans: [c]	18. Ans: [d]	19. Ans: [b]	20. Ans: [b]
8.4571	2	712	10	1	1