CSC4008 Assignment 5: SVM

Name: 费祥　　Student ID: 120090414

T1. The prime problem:

$$\min_{w,b} \frac{1}{2}\|w\|^2$$

$$\text{s.t. } 1-y_i(w^T x_i+b)\leq 0, \forall i$$

Lagrange function: $L(w,b,\alpha)=\frac{1}{2}\|w\|^2+\sum_i^m \alpha_i(1-y_i(w^Tx_i+b))$ and $\alpha_i \geq 0, \forall i$

KKT condition:

stationarity: $\frac{\partial L}{\partial w}=0 \Rightarrow w=\sum_i^m \alpha_i y_i x_i$

$\frac{\partial L}{\partial b}=0 \Rightarrow \sum_i^m \alpha_i y_i=0$

Feasibility: $\alpha_i \geq 0$, $1-y_i(w^Tx_i+b)\leq 0, \forall i$

Complementary slackness: $\alpha_i(1-y_i(w^Tx_i+b))=0, \forall i$

corresponding dual problem:

$$\max_{\alpha} \sum_i^m \alpha_i-\frac{1}{2}\sum_{i,j}^m \alpha_i\alpha_j y_i y_j x_i^T x_j$$

$$\text{s.t. } \sum_i^m \alpha_i y_i=0$$

and the prime solution: $w=\sum_i^m \alpha_i y_i x_i$

$b=y_j-\sum_i^m \alpha_i y_i x_i^T x_j$, $\forall j \in S$, S is the support set.

(1) $\max_\alpha \sum_i^3 \alpha_i-\frac{1}{2}\sum_{i,j}^3 \alpha_i\alpha_j y_i y_j x_i^T x_j$

$=\max_\alpha \alpha_1+\alpha_2+\alpha_3-\frac{1}{2}(2\alpha_1^2+\alpha_1\alpha_3+2\alpha_2^2+\alpha_2\alpha_3+\alpha_3^2+\alpha_1\alpha_3+\alpha_2\alpha_3)$

$=\max_\alpha \alpha_1+\alpha_2+\alpha_3-\alpha_1^2-\alpha_2^2-\frac{1}{2}\alpha_3^2-\alpha_1\alpha_3-\alpha_2\alpha_3$

s.t. $\alpha_1+\alpha_2-\alpha_3=0$

$\therefore \max_\alpha g(\alpha)=\max_\alpha 2\alpha_1+2\alpha_2-\frac{5}{2}\alpha_1^2-\frac{5}{2}\alpha_2^2-3\alpha_1\alpha_2$

$\therefore \frac{\partial g}{\partial \alpha_1}=2-5\alpha_1-3\alpha_2=0$, $\frac{\partial g}{\partial \alpha_2}=2-5\alpha_2-3\alpha_1=0$

$\therefore \alpha_1 = \alpha_2 = \frac{1}{4}$, $\alpha_3 = \frac{1}{2}$

$\therefore w = \frac{1}{4} \times 1 \times \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \frac{1}{4} \times 1 \times \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \frac{1}{2} \times (-1) \times \begin{bmatrix} 0 \\ -1 \end{bmatrix}$

$= \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$b = 1 - \left(\frac{1}{4} \times 1 \times 2 + \frac{1}{4} \times 1 \times 0 + \frac{1}{2} \times (-1) \times (-1)\right) = 0$

(2) $\max_{\alpha} \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - \frac{1}{2}(2\alpha_1^2 + \alpha_1\alpha_3 + 2\alpha_1\alpha_4 + 2\alpha_2^2 + \alpha_2\alpha_3 + 2\alpha_2\alpha_4$
$+ \alpha_1\alpha_3 + \alpha_2\alpha_3 + \alpha_3^2 + 2\alpha_3\alpha_4 + 2\alpha_1\alpha_4 + 2\alpha_2\alpha_4$
$+ 2\alpha_3\alpha_4 + 4\alpha_4^2)$

$= \max_{\alpha} \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - \alpha_1^2 - \alpha_2^2 - \frac{1}{2}\alpha_3^2 - 2\alpha_4^2 - \alpha_1\alpha_3 - \alpha_2\alpha_3 - 2\alpha_1\alpha_4$
$-2\alpha_2\alpha_4 - 2\alpha_3\alpha_4 = \max_{\alpha} g(\alpha)$

s.t. $\alpha_1 + \alpha_2 - \alpha_3 - \alpha_4 = 0$

If $\alpha_1 > 0, \alpha_2 > 0, \alpha_3 > 0, \alpha_4 > 0$

$\frac{\partial g}{\partial \alpha_1} = 1 - 2\alpha_1 - \alpha_3 - 2\alpha_4 = 0$

$\frac{\partial g}{\partial \alpha_2} = 1 - 2\alpha_2 - \alpha_3 - 2\alpha_4 = 0$

$\frac{\partial g}{\partial \alpha_3} = 1 - \alpha_3 - \alpha_1 - \alpha_2 - 2\alpha_4 = 0$

$\frac{\partial g}{\partial \alpha_4} = 1 - 4\alpha_4 - 2\alpha_1 - 2\alpha_2 - 2\alpha_3 = 0$

contradiction, then we find $\alpha_4 = 0$, margin is the biggest

$\therefore \alpha_1 = \alpha_2 = \frac{1}{4}$, $\alpha_3 = \frac{1}{2}$, $\alpha_4 = 0$

$\therefore w = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $b = 0$

T3、prime problem:

$$\min_{w,b,\varepsilon} \frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^{N}\varepsilon_i\right)^2$$

$$\text{s.t. } 1-\varepsilon_i - y_i(w^T x_i + b) \le 0,\ -\varepsilon_i \le 0,\ \forall i$$

Lagrange function:

$$L(w,b,\varepsilon,\alpha,\mu) = \frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^{N}\varepsilon_i\right)^2 + \sum_{i=1}^{N}\left[\alpha_i\left(1-\varepsilon_i - y_i(w^T x_i + b)\right) - \mu_i\varepsilon_i\right]$$

and $\alpha_i, \mu_i \geq 0,\ \forall i$

KKT conditions:

Stationarity: $\frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{N}\alpha_i y_i x_i$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i}^{m}\alpha_i y_i = 0$$

$$\frac{\partial L}{\partial \varepsilon_i} = 0 \Rightarrow C\cdot 2\sum_{i=1}^{N}\varepsilon_i - \alpha_i - \mu_i = 0$$

Feasibility: $\alpha_i \geq 0,\ 1-\varepsilon_i - y_i(w^T x_i + b) \le 0,\ \varepsilon_i \geq 0,\ \mu_i \geq 0,\ \forall i$

Complementary slackness:

$$\alpha_i\left(1-\varepsilon_i - y_i(w^T x_i + b)\right) = 0,\ \mu_i\varepsilon_i = 0,\ \forall i$$

replacing all KKT conditions into Lagrange function to eliminate primal variables, we have:

$$L(\alpha,\mu) = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1,j=1}^{N}\alpha_i\alpha_j y_i y_j x_i^T x_j - \frac{(\alpha_i+\mu_i)^2}{4C}$$

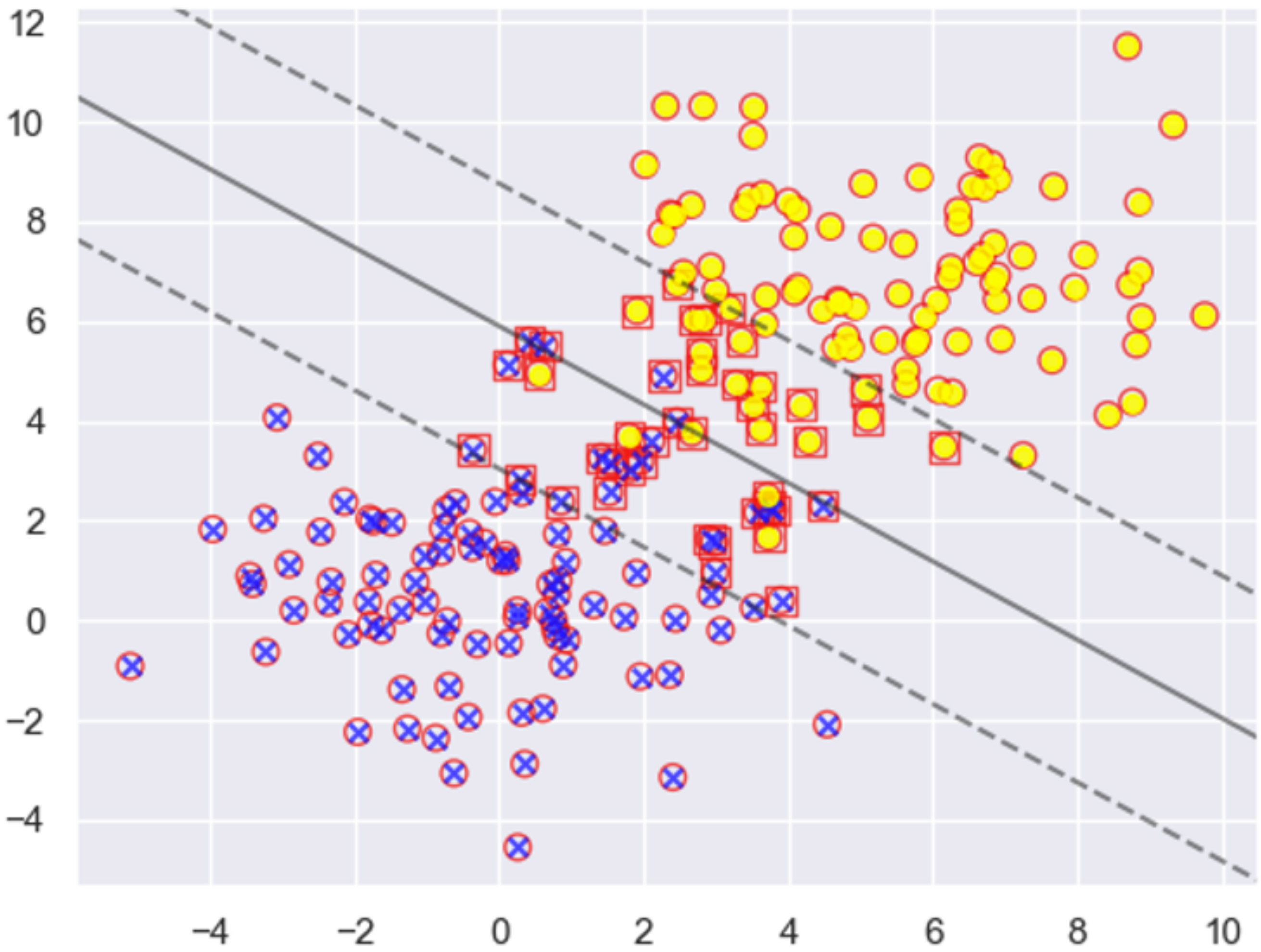
dual problem:

$$\max_{\alpha} \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1,j=1}^{N}\alpha_i\alpha_j y_i y_j x_i^T x_j - \frac{(\alpha_i+\mu_i)^2}{4C}$$

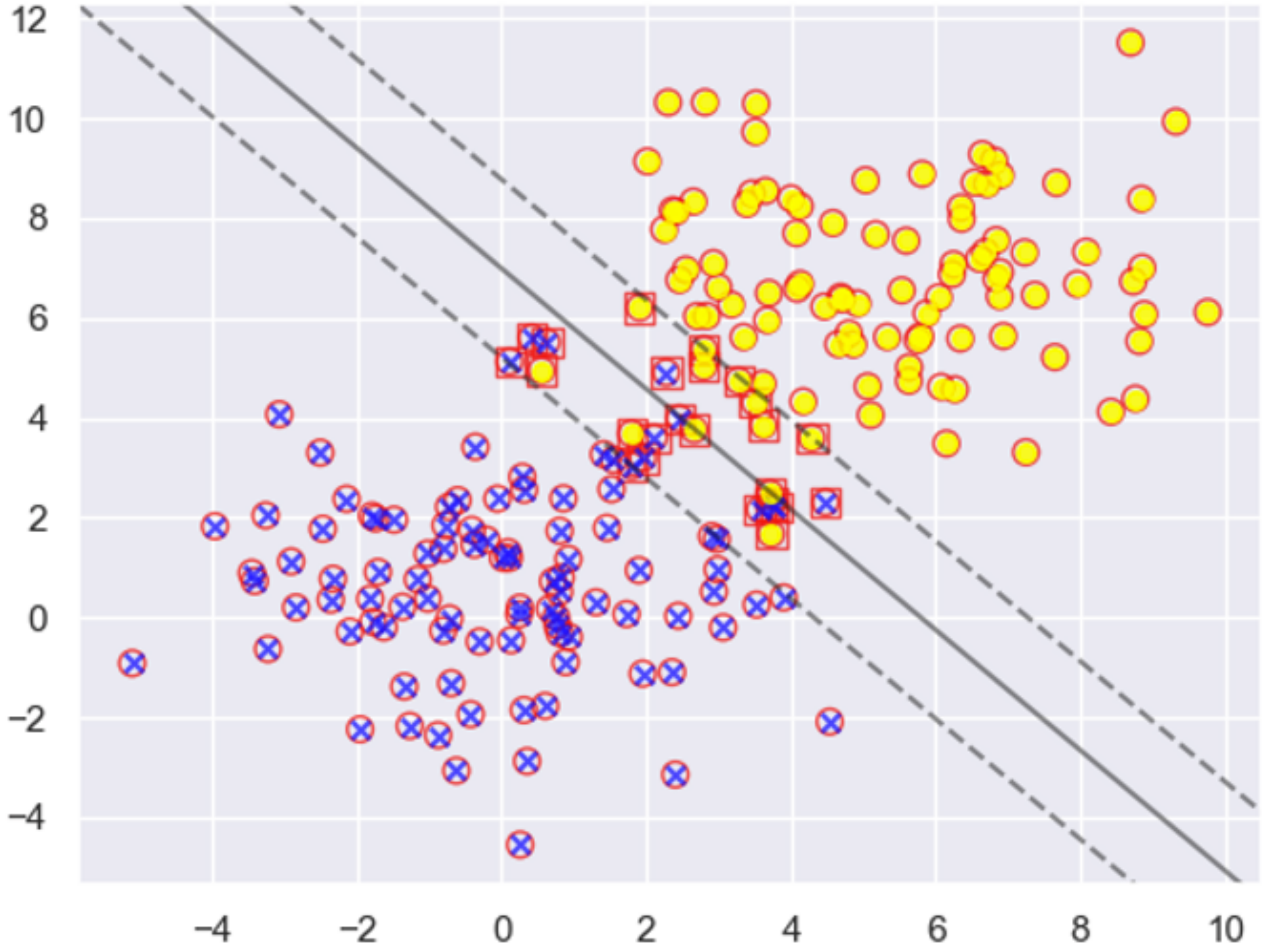
$$\text{s.t. } \sum_{i=1}^{N}\alpha_i y_i = 0,\ \alpha_i \geq 0,\ \mu_i \geq 0$$

T2: 费祥(120090414)，沈嘉佑(119010259)

when C = 0.01:



when C = 0.1:



$\alpha_i = 0$: means the corresponding data are correctly classified and doesn't contribute to the classifier, locating outside of the margin. The data is not support vector.

$\alpha_i = C$: in this case, μi = 0; then we have ξi > 0. The corresponding data contributes to the classifier, locating inside the margin, here we should pay attention that these points are not exactly on the margin. In fact, the points locate on the margin satisfies $0 < \alpha_i < C$.

ξi > 0 data points are those inside the margin, the corresponding $\alpha_i = C$. They are a part of support vectors.