Homework

Question 1

(a)

If the above set of constraints (second line of equations above) is satisfied, all the training data will be correctly classified.

(b)

The Lagrangian function is:

$$L(\underline{w}, w_0, \lambda) = \frac{1}{2}||\underline{w}||^2 + \sum_{i=1}^{N} \lambda_i (1 - z_i(\underline{w}^T \underline{u}_i + w_0)) \tag{1}$$

where $\lambda_i \geq 0, i = 1, 2, 3, \cdots, N$

KKT conditions:

$$\begin{aligned} \lambda_i &\geq 0 \\ z_i(\underline{w}^T \underline{u}_i + w_0) - 1 &\geq 0 \\ \lambda_i(1 - z_i(\underline{w}^T \underline{u}_i + w_0)) &= 0 \end{aligned} \tag{2}$$

(c)

Let the derivation of L w.r.t $\underline{w}, w_0$ to zero, we can get that

$$\begin{aligned} \frac{\partial L}{\partial \underline{w}} &= \underline{w} - \sum_{i=1}^{N} \lambda_i z_i \underline{u}_i = 0 \\ \underline{w}^* &= \sum_{i=1}^{N} \lambda_i z_i \underline{u}_i \\ \frac{\partial L}{\partial w_0} &= \sum_{i=1}^{N} \lambda_i z_i = 0 \end{aligned} \tag{3}$$

replace $\underline{w}, w_0$ in L, we can obtain the dual problem

$$\begin{aligned} &\max_{\lambda_i} L_D(\underline{\lambda}) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j z_i z_j \underline{u}_i^T \underline{u}_j \\ &s.t. \sum_{i=1}^{N} \lambda_i z_i = 0 \end{aligned} \tag{4}$$

The other two KKT conditions are

$$\begin{aligned} \lambda_i &\geq 0 \\ z_i(\underline{w}^T \underline{u}_i + w_0) - 1 &\geq 0 \end{aligned} \tag{5}$$

Question 2

(a)

From the question 1, we can get that

$$L_D^{'}(\underline{\lambda}, \mu) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j z_i z_j \underline{u}_i^T \underline{u}_j + \mu(\sum_{i=1}^{N} z_i \lambda_i) \tag{6}$$

Let the derivation of $L_D^{'}$ w.r.t $\underline{\lambda}, \mu$ to zero, we can get that

$$\begin{aligned} \frac{\partial L_D^{'}}{\partial \underline{\lambda}} &= 1 + \mu z_i - \sum_{j=1}^{N} \lambda_j z_i z_j \underline{u}_i^T \underline{u}_j = 0 \\ \frac{\partial L_D^{'}}{\partial \mu} &= \sum_{i=1}^{N} z_i \lambda_i = 0 \end{aligned} \tag{7}$$

The final form of $\underline{\underline{A}}$, and $\underline{b}$ are :

$$\underline{\underline{A}} = \begin{bmatrix} z_1^2 \underline{u}_1^T \underline{u}_1 & z_1 z_2 \underline{u}_1^T \underline{u}_2 & z_1 z_3 \underline{u}_1^T \underline{u}_3 & -z_1 \\ z_2 z_1 \underline{u}_2^T \underline{u}_1 & z_2^2 \underline{u}_2^T \underline{u}_2 & z_2 z_3 \underline{u}_2^T \underline{u}_3 & -z_2 \\ z_3 z_1 \underline{u}_3^T \underline{u}_1 & z_3 z_2 \underline{u}_3^T \underline{u}_2 & z_3^2 \underline{u}_3^T \underline{u}_3 & -z_3 \\ z_1 & z_2 & z_3 & 0 \end{bmatrix} \tag{8}$$

$$\underline{b} = [1 \quad 1 \quad 1 \quad 0]^T \tag{9}$$

(c)

λ is [0.222, 0.222, 0.444], μ is 1

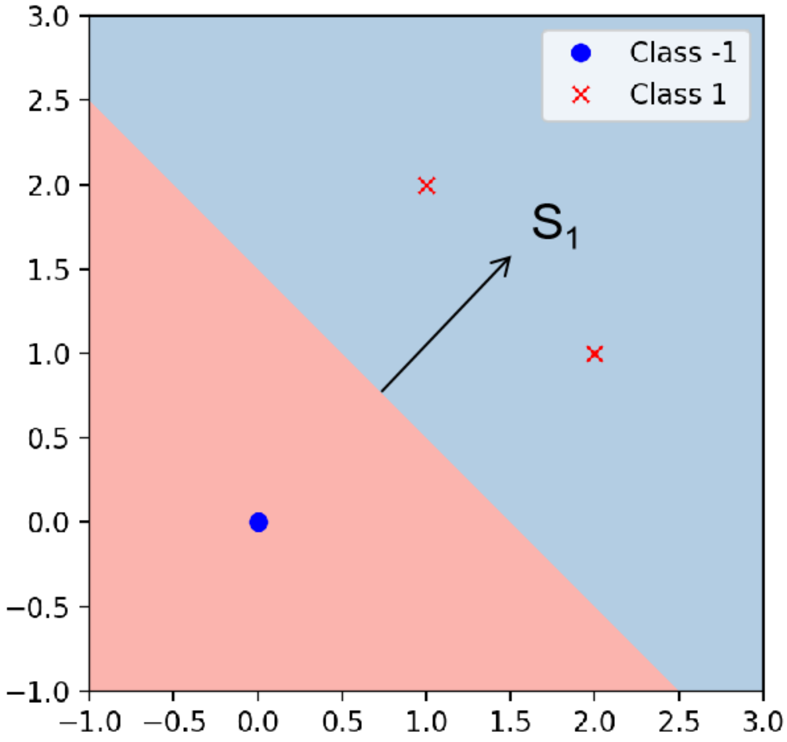
KKT conditions 1 is satisfied.

KKT conditions 2 is satisfied.

w is [0.667, 0.667], w_0 is -1

KKT conditions 3 is satisfied.

(d)



(e)

From the figure above, we can find that the decision boundary correctly classify the training data. It is the maximum-margin boundary that correctly classified the data.

First, we can compute the two class' distance by u_1, u_2, u_3 is $D(c_{+1}, c_{-1}) = \frac{|3|}{\sqrt{1^2+1^2}} = \frac{3\sqrt{2}}{2}$, and the u_3 to the decision boundary's distance is $D(c_{-1}, d) = \frac{|\frac{3}{2}|}{\sqrt{1^2+1^2}} = \frac{3\sqrt{2}}{4}$.

Thus, the distance from u_3 to the decision boundary is just the half between the two class' distance. The decision boundary is the maximum-margin boundary that correctly classified the data.

(f)

(i)

λ is [2, 2, 4], μ is 5

KKT conditions 1 is satisfied.

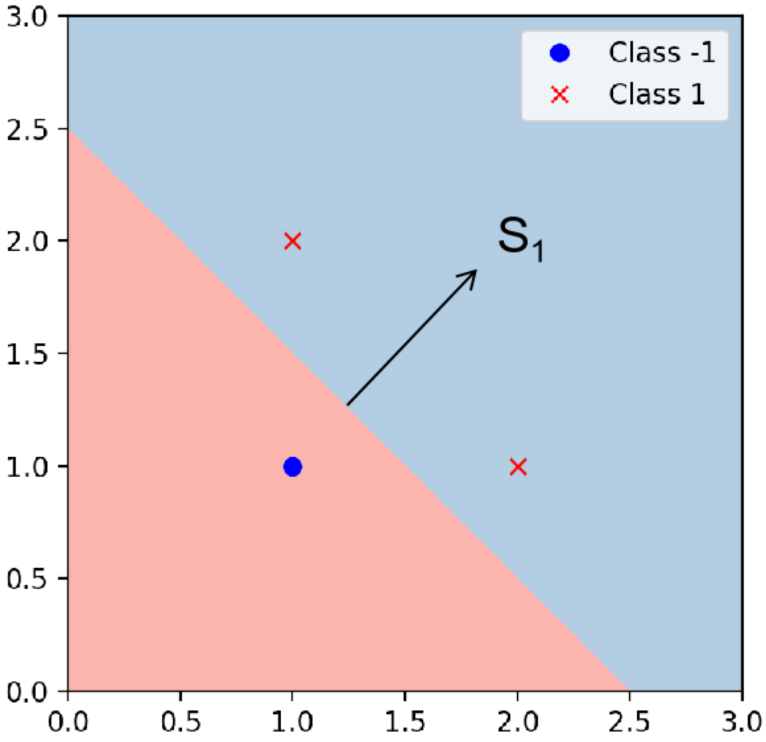
KKT conditions 2 is satisfied.

w is [2, 2], w_0 is -5

KKT conditions 3 is satisfied.

(ii)

The decision boundary is shown in the figure below:



(iii)

From the figure above, we can find that the decision boundary correctly classify the training data. It is the maximum-margin boundary that correctly classified the data.

First, we can compute the two class' distance is $D(c_{+1}, c_{-1}) = \frac{|1|}{\sqrt{1^2+1^2}} = \frac{\sqrt{2}}{2}$, and the u_3 to the decision boundary's distance is $D(c_{-1}, d) = \frac{|\frac{1}{2}|}{\sqrt{1^2+1^2}} = \frac{\sqrt{2}}{4}$.

Thus, the distance from u_3 to the decision boundary is just the half between the two class' distance. The decision boundary is the maximum-margin boundary that correctly classified the data.

(iv)

Compare with (d), we can find that the decision boundary is more close to the two class, this is because u_3 is more close to the class +1. But the direction of the decision boundary is not changed.

(g)

(i)

The decision boundary's direction will change for the reason that u_3 has changed to the different direction to the class +1.

(ii)

λ is [1.6, 0, 1.6], μ is 2.2

KKT conditions 1 is satisfied.

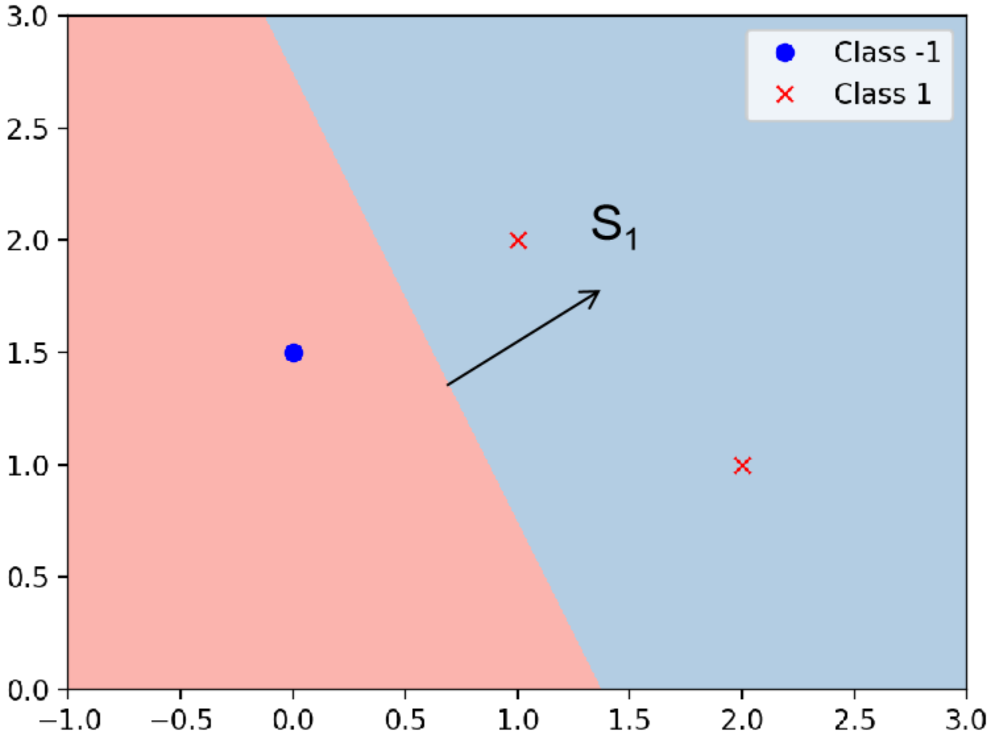
KKT conditions 2 is satisfied.

w is [1.6, 0.8], w_0 is -2.2

KKT conditions 3 is satisfied.

(iii)

The decision boundary is shown in the figure below:



(iv)

From the figure above, we can find that the decision boundary correctly classify the training data. And it is the maximum-margin boundary that correctly classified the data.

First, we can compute the two class' distance with support vector u_1, u_3 is $D(c_{+1}, c_{-1}) = \sqrt{1^2 + 0.5^2} = \frac{\sqrt{5}}{2}$, and the u_3 to the decision boundary's distance is $D(c_{-1}, d) = \frac{|\frac{5}{4}|}{\sqrt{1^2+2^2}} = \frac{\sqrt{5}}{4}$.

Thus, the distance from u_3 to the decision boundary is just the half between the two class' distance. The decision boundary is the maximum-margin boundary that correctly classified the data.

(v)

Compared with (d) and (f), we can find the decision boundary's direction has changed due to the point u_3's change.