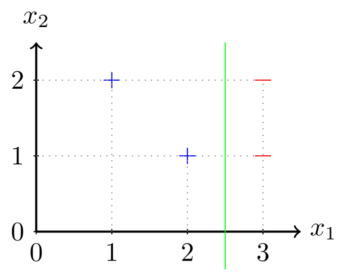
CS680, Spring 2020, Assignment 4

Zhijie Wang, [redacted]

June 5, 2020

Exercise 1

1. The optimal solution should maximize the minimum distance from data points to the hyperplane. In this problem the data points are 2d and therefore the solution should be a line separate two classes of points. Suppose the two dimension are x_1 and x_2. The closest positive and negative points are (2,1) and (3,1), thus, the optimal solution is the line which perpendicular to the line connected (2,1) and (3,1), $x_1 \quad 2.5$. Hence, $\mathbf{w}* \quad [-1, 0]^T, b* \quad 2.5$



2. Use the slack variables ξ, we can formulate the optimization problem as

$$\begin{aligned} \min_{\mathbf{w},b,\xi} \quad & \frac{1}{2}\|\mathbf{w}\|_2^2 + C\sum_{i=1}^{n}\xi^2 \\ \text{s.t.} \quad & 1 - y_i(\mathbf{w}^T\mathbf{x}_i + b) \leqslant \xi_i, i \quad 1, 2, \ldots, n \end{aligned} \tag{1}$$

Then the Lagrangian is

$$\min_{\mathbf{w},b,\xi}\max_{\alpha\geqslant 0}\frac{1}{2}\|\mathbf{w}\|_2^2 + \sum_{i=1}^{n} C\xi_i^2 + \alpha_i(1 - \xi_i - y_i(\mathbf{w}^T\mathbf{x}_i + b)) \tag{2}$$

Take the derivative of Lagrangian, and let them equal to zero, we can obtain

$$\begin{aligned} \frac{\partial}{\partial \xi_i} \quad & 2C\xi_i - \alpha_i \quad 0 \Rightarrow \xi_i \quad \frac{1}{2C}\alpha_i \\ \frac{\partial}{\partial \mathbf{w}} \quad & \mathbf{w} - \sum_{i=1}^{n}\alpha_i y_i \mathbf{x}_i \quad 0 \Rightarrow \mathbf{w} \quad \sum_{i=1}^{n}\alpha_i y_i \mathbf{x}_i \\ \frac{\partial}{\partial b} \quad & \sum_{i=1}^{n}\alpha_i y_i \quad 0 \end{aligned} \tag{3}$$

Put these back into the Lagrangian, we can obtain

$$\begin{aligned} & \frac{1}{2}\|\mathbf{w}\|_2^2 + \sum_{i=1}^{n} C\xi_i^2 + \alpha_i(1 - \xi_i - y_i(\mathbf{w}^T\mathbf{x}_i + b)) \\ & \frac{1}{2}\|\sum_{i=1}^{n}\alpha_i y_i \mathbf{x}_i\|_2^2 + \sum_{i=1}^{n}\alpha_i + \sum_{i=1}^{n}\frac{1}{4C}\alpha_i^2 - \sum_{i=1}^{n}\frac{1}{2C}\alpha_i^2 - \sum_{i=1}^{n}\alpha_i y_i(\sum_{j=1}^{n}\alpha_j y_j \mathbf{x}_j)^T\mathbf{x}_i \\ & \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j \mathbf{x}_i^T\mathbf{x}_j - \frac{1}{4C}\sum_{i=1}^{n}\alpha_i^2 \end{aligned} \tag{4}$$

Finally we can get,

$$
\begin{aligned}
\max_{\alpha \geqslant 0} \quad & \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j - \frac{1}{4C} \sum_{i=1}^{n} \alpha_i^2 \\
\text{s.t.} \quad & \sum_{i=1}^{n} \alpha_i y_i \quad 0
\end{aligned}
\tag{5}
$$

3.

$$
\begin{aligned}
& \frac{\partial}{\partial \mathbf{w}} C \max\{1 - y_i(\mathbf{w}^T \mathbf{x}_i + b), 0\}^2 \\
& \begin{cases} \frac{\partial}{\partial \mathbf{w}} C(1 - y_i(\mathbf{w}^T \mathbf{x}_i + b))^2 & 1 - y_i(\mathbf{w}^T \mathbf{x}_i + b) > 0 \\ 0 & \text{otherwise} \end{cases} \\
& \begin{cases} -2C \cdot y_i(1 - y_i(\mathbf{w}^T \mathbf{x}_i + b))\mathbf{x}_i & 1 - y_i(\mathbf{w}^T \mathbf{x}_i + b) > 0 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
\tag{6}
$$

$$
\begin{aligned}
& \frac{\partial}{\partial b} C \max\{1 - y_i(\mathbf{w}^T \mathbf{x}_i + b), 0\}^2 \\
& \begin{cases} \frac{\partial}{\partial b} C(1 - y_i(\mathbf{w}^T \mathbf{x}_i + b))^2 & 1 - y_i(\mathbf{w}^T \mathbf{x}_i + b) > 0 \\ 0 & \text{otherwise} \end{cases} \\
& \begin{cases} -2C \cdot y_i(1 - y_i(\mathbf{w}^T \mathbf{x}_i + b)) & 1 - y_i(\mathbf{w}^T \mathbf{x}_i + b) > 0 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
\tag{7}
$$

4. First expand the expression of (3) in the Problem.

$$
\begin{aligned}
& \frac{1}{2\eta} \|\mathbf{z} - \mathbf{w}\|_2^2 + \frac{1}{2} \|\mathbf{z}\|_2^2 \\
& \frac{1}{2\eta} ((\mathbf{z} - \mathbf{w})^T (\mathbf{z} - \mathbf{w})) + \frac{1}{2} \mathbf{z}^T \mathbf{z} \\
& \frac{1}{2\eta} (\mathbf{z}^T \mathbf{z} - \mathbf{w}^T \mathbf{z} - \mathbf{z}^T \mathbf{w} + \mathbf{w}^T \mathbf{w}) + \frac{1}{2} \mathbf{z}^T \mathbf{z}
\end{aligned}
\tag{8}
$$

Take the derivative of above equation and let it go to zero.

$$
\begin{aligned}
\frac{1}{\eta} \mathbf{z} - \frac{1}{2\eta} \mathbf{w} - \frac{1}{2\eta} \mathbf{w} + \mathbf{z} & \quad 0 \\
(\frac{1}{\eta} + 1) \mathbf{z} & \quad \frac{1}{\eta} \mathbf{w} \\
\mathbf{z} & \quad \frac{1}{1 + \eta} \mathbf{w}
\end{aligned}
\tag{9}
$$

Therefore, $\mathsf{P}^{\eta}(\mathbf{w}) \quad \frac{1}{1+\eta} \mathbf{w}$

5. cs680-asg4.py

6. When C equals to 100, step size η is 0.001, we can obtain $w \quad [-1.992000, -0.003962]^T$, $b \quad 4.948301$. Since we can multiply each term of $\mathbf{w}^*$ and b^* in Question 1 with 2, we can find that the result is close to the optimal solution we've derived in Question 1.