CSE 5526 Handout:
Linear SVM for a Linearly Inseparable Problem

Problem: For the following linearly inseparable problem:

$$\mathbf{x}_1 = (0,0)^T \quad d_1 = -1$$

$$\mathbf{x}_2 = (0,2)^T \quad d_2 = 1$$

$$\mathbf{x}_3 = (2,0)^T \quad d_3 = -1$$

$$\mathbf{x}_4 = (2,2)^T \quad d_4 = 1$$

$$\mathbf{x}_5 = (0,3)^T \quad d_5 = -1$$

Find the optimal hyperplane with $C = 2$.

Solution: For this classification problem, we can use a soft margin linear SVM to give a solution. The objective function for the dual problem is:

$$\begin{aligned} Q(\boldsymbol{\alpha}) = & \textstyle\sum_{i=1}^{5} \alpha_i - \frac{1}{2}\left(\sum_{i=1}^{5} \sum_{j=1}^{5} \alpha_i \alpha_j d_i d_j \mathbf{x}_i^T \mathbf{x}_j\right) \\ = & \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 \\ & -\tfrac{1}{2}(4\alpha_2^2 + 8\alpha_2\alpha_4 - 12\alpha_2\alpha_5 + 4\alpha_3^2 - 8\alpha_3\alpha_4 + 8\alpha_4^2 - 12\alpha_4\alpha_5 + 9\alpha_5^2) \end{aligned} \quad (1)$$

subject to the constraints

$$\text{(a).} \quad \sum_{i=1}^{5} \alpha_i d_i = -\alpha_1 + \alpha_2 - \alpha_3 + \alpha_4 - \alpha_5 = 0$$

$$\text{(b).} \quad 0 \le \alpha_i \le C \quad \text{for} \quad i = 1,2,\dots,5$$

Note that, this optimization problem has a more stringent constraint $0 \le \alpha_i \le C$ than in a linearly separable case, which makes the optimization more difficult. For this specific problem, we can directly solve it. By using the first constraint, the dual problem can be rewritten as

$$Q(\boldsymbol{\alpha}) = 2\alpha_2 + 2\alpha_4 - \tfrac{1}{2}(4\alpha_2^2 + 8\alpha_2\alpha_4 - 12\alpha_2\alpha_5 + 4\alpha_3^2 - 8\alpha_3\alpha_4 + 8\alpha_4^2 - 12\alpha_4\alpha_5 + 9\alpha_5^2) \quad (2)$$

Calculating the partial derivative of Q with respect to each Lagrange multiplier yields the following set of simultaneous equations:

$$
\begin{cases}
\frac{\partial Q}{\partial \alpha_2} = 2 - \frac{1}{2}[4(2\alpha_2 + 2\alpha_4) - 12\alpha_5] = 0 \\
\frac{\partial Q}{\partial \alpha_3} = -\frac{1}{2}[4(2\alpha_3 - 2\alpha_4)] = 0 \\
\frac{\partial Q}{\partial \alpha_4} = 2 - \frac{1}{2}[4(2\alpha_2 + 2\alpha_4) - 4(2\alpha_3 - 2\alpha_4) - 12\alpha_5] = 0 \\
\frac{\partial Q}{\partial \alpha_5} = -\frac{1}{2}(-12\alpha_2 - 12\alpha_4 + 18\alpha_5) = 0
\end{cases} \tag{3}
$$

$$
\Rightarrow \begin{cases}
2\alpha_2 + 2\alpha_4 - 3\alpha_5 = 1 \\
\alpha_3 = \alpha_4 \\
2\alpha_2 - 2\alpha_3 + 2\alpha_4 - 3\alpha_5 = 1 \\
2\alpha_2 + 2\alpha_4 - 3\alpha_5 = 0
\end{cases} \tag{4}
$$

This set of equations is under-constrained. Therefore, the maximum of Q must be located on the boundary for some multipliers, i.e., $\exists i$ s.t. $\alpha_i = 0$ or $\alpha_i = C$. We can use some tricks to get the optimal values. By (2), we have

$$
\begin{aligned}
Q = & \quad 2\alpha_2 + 2\alpha_4 - \frac{1}{2}[(2\alpha_3 - 2\alpha_4)^2 + (2\alpha_2 - 2\alpha_5)^2 + 4\alpha_4^2 + 5\alpha_5^2 - 4\alpha_2\alpha_5 + 8\alpha_2\alpha_4 - 12\alpha_4\alpha_5] \\
= & \quad 2\alpha_2 + 2\alpha_4 - \frac{1}{2}[(2\alpha_3 - 2\alpha_4)^2 + (2\alpha_2 - 2\alpha_5)^2 + (2\alpha_4 - \alpha_5)^2 + 4(\alpha_2 - \alpha_5)(2\alpha_4 - \alpha_5)]
\end{aligned} \tag{5}
$$

Let $\beta_1 = 2\alpha_3 - 2\alpha_4$, $\beta_2 = 2\alpha_2 - 2\alpha_5$ and $\beta_3 = 2\alpha_4 - \alpha_5$ and substitute them into Eq. (5). We then have

$$
\begin{aligned}
Q = & \quad 3\alpha_5 + \beta_2 + \beta_3 - \frac{1}{2}(\beta_1^2 + \beta_2^2 + \beta_3^2 + 2\beta_2\beta_3) \\
= & \quad 3\alpha_5 - \frac{1}{2}(\beta_1^2 + \beta_2^2 + \beta_3^2 + 2\beta_2\beta_3 - 2\beta_2 - 2\beta_3) \\
= & \quad 3\alpha_5 + \frac{1}{2} - \frac{1}{2}[\beta_1^2 + (\beta_2 + \beta_3 - 1)^2]
\end{aligned} \tag{6}
$$

In order to maximize Q, we have $\alpha_5 = C$, $\beta_1 = 0$ and $\beta_2 + \beta_3 - 1 = 0$. So

$$
\alpha_5 = C, \alpha_3 = \alpha_4, \alpha_2 + \alpha_4 = \frac{3C + 1}{2}
$$

Since $\alpha_1 = \alpha_2 - \alpha_3 + \alpha_4 - \alpha_5 \geq 0$ and $0 \leq \alpha_i \leq C$, we have $\alpha_2 - \alpha_5 \geq 0 \Rightarrow \alpha_2 = C$

Therefore we have

$$\Rightarrow \begin{cases} \alpha_2 = C \\ \alpha_5 = C \\ \alpha_3 = \frac{C+1}{2} \\ \alpha_4 = \frac{C+1}{2} \\ \alpha_1 = 0 \text{ (from Constraint (a))} \end{cases} \tag{7}$$

Note that $0 \leq \alpha_i \leq C$ is satisfied when $C \geq 1$. So, with $C = 2$, we have

$$\alpha_1 = 0, \alpha_2 = 2, \alpha_3 = \frac{3}{2}, \alpha_4 = \frac{3}{2}, \alpha_5 = 2 \tag{8}$$

Consequently, the optimal weight vector is:

$$\mathbf{w} = \sum_{i=1}^{5} \alpha_i d_i \mathbf{x}_i = (0,1)^T$$

To compute the optimal bias b, we use the obtained $\mathbf{w}$ and the following equation:

$$d_i^{(s)}\left(\mathbf{w}^T \mathbf{x}_i^{(s)} + b\right) = 1 - \xi_i \tag{9}$$

where, $\mathbf{x}_i^{(s)}$ and $d_i^{(s)}$ denote a support vector and its desired output, respectively.

Note that, for a support vector with $0 < \alpha_i < C$, we have $\xi_i = 0$. In this case, $\mathbf{x}_3$ is such a support vector so it is easy to calculate that $b = -1$.

Finally, the optimal decision boundary is:

$$\mathbf{w}^T \mathbf{x} + b = 0 \Rightarrow x_2 = 1$$

Figure 1 shows the data points and the decision boundary. Note that, as we use the linear SVM to solve the linearly inseparable problem, we are not able to perfectly classify the data.

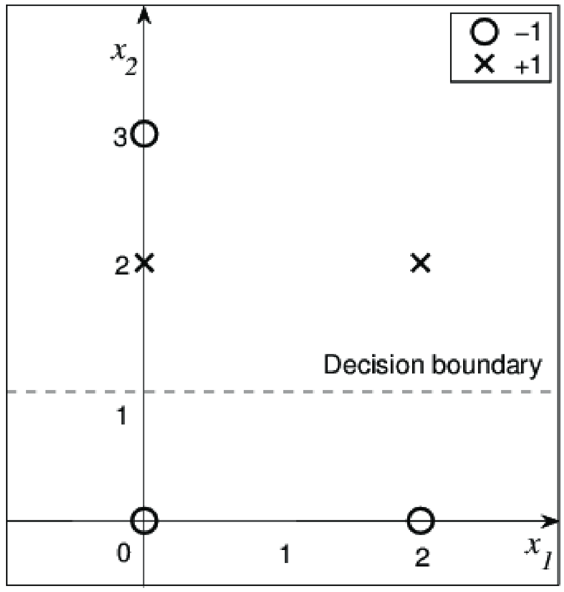

Figure 1