Example 3.5

Consider the two-class problem where class ω_1 (+1) consists of the vectors $\boldsymbol{x}_1 = [-1,\ 1]^T$, $\boldsymbol{x}_2 = [-1,\ -1]^T$, while class ω_2 (−1) consists of the vectors $\boldsymbol{x}_3 = [1,\ -1]^T$, $\boldsymbol{x}_4 = [1,\ 1]^T$.

We will demonstrate how the utilization of the SVM approach leads to the optimal separating hyperplane, which is $x_1 = 0$ and, in addtion, that this is obtained for different sets of Lagrange multipliers. In the sequel, we solve the problem in its Wolfe dual representation, i.e.,

$$\max_{\lambda}\left(\sum_{i=1}^{N}\lambda_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j\right) \equiv J_1^*(\boldsymbol{x}) \tag{1}$$

subject to the constrains

$$\sum_{i=1}^{N}\lambda_i y_i = 0 \tag{2}$$

$$\lambda_i \geq 0, \qquad i = 1, \ldots, N \tag{3}$$

where, in our case it is $N = 4$, $y_1 = y_2 = +1$ and $y_3 = y_4 = -1$. In the sequel, for notational simplicity, we write J instead of $J_1^*(\boldsymbol{x})$.

It is straightforward to deduce that $y_1y_1\boldsymbol{x}_1^T\boldsymbol{x}_1 = 2$, $y_1y_2\boldsymbol{x}_1^T\boldsymbol{x}_2 = 0$, $y_1y_3\boldsymbol{x}_1^T\boldsymbol{x}_3 = 2$, $y_1y_4\boldsymbol{x}_1^T\boldsymbol{x}_4 = 0$, $y_2y_2\boldsymbol{x}_2^T\boldsymbol{x}_2 = 2$, $y_2y_3\boldsymbol{x}_2^T\boldsymbol{x}_3 = 0$, $y_2y_4\boldsymbol{x}_2^T\boldsymbol{x}_4 = 2$, $y_3y_3\boldsymbol{x}_3^T\boldsymbol{x}_3 = 2$, $y_3y_4\boldsymbol{x}_3^T\boldsymbol{x}_4 = 0$, $y_4y_4\boldsymbol{x}_4^T\boldsymbol{x}_4 = 2$. Based on these results, J from (1) becomes

$$J = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 - \lambda_1^2 - \lambda_2^2 - \lambda_3^2 - \lambda_4^2 - 2\lambda_1\lambda_3 - 2\lambda_2\lambda_4 \tag{4}$$

Taking the gradient of J with respect to λ_i, $i = 1, \ldots, 4$, we have

$$\frac{\partial J}{\partial \lambda_1} = 1 - 2\lambda_1 - 2\lambda_3$$

$$\frac{\partial J}{\partial \lambda_2} = 1 - 2\lambda_2 - 2\lambda_4$$

$$\frac{\partial J}{\partial \lambda_3} = 1 - 2\lambda_1 - 2\lambda_3$$
$$\frac{\partial J}{\partial \lambda_4} = 1 - 2\lambda_2 - 2\lambda_4$$

Setting the gradients equal to 0 and after a bit of algebra, we end up with the following two independent equations

$$\lambda_1 + \lambda_3 = \frac{1}{2} \tag{5}$$

$$\lambda_2 + \lambda_4 = \frac{1}{2} \tag{6}$$

In addition, eq. (2) gives

$$\lambda_1 + \lambda_2 = \lambda_3 + \lambda_4 \tag{7}$$

Solving eqs. (5) and (6) with respect to λ_3 and λ_2, respectively, and substituting to (7), we conclude, after some algebra, that $\lambda_1 = \lambda_4$ and, as a consequence, $\lambda_2 = \lambda_3$. Taking into account that all λ_i's are nonnegative, we have the following set of relations for them

$$\lambda_1 = \lambda_4 \tag{8}$$

$$\lambda_2 = \lambda_3 \tag{9}$$

$$0 \leq \lambda_i \leq \frac{1}{2}, \qquad i = 1, 2, 3, 4 \tag{10}$$

Let $u \in [0,\ \frac{1}{2}]$. Setting $\lambda_1 = \lambda_4 = u$, which in turn implies that $\lambda_2 = \lambda_3 = \frac{1}{2} - u$, $\boldsymbol{w}$ is computed as follows:

$$\boldsymbol{w} = u \begin{bmatrix} -1 \\ 1 \end{bmatrix} + (\frac{1}{2} - u) \begin{bmatrix} -1 \\ -1 \end{bmatrix} - (\frac{1}{2} - u) \begin{bmatrix} 1 \\ -1 \end{bmatrix} - u \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \tag{11}$$

Observe that, although there are more than one valid set of values for the λ_i's, all of them lead to the same solution for $\boldsymbol{w}$.

In addition, w_0 can be implicitly obtained via the equations

$$\lambda_i[y_i(\boldsymbol{w}^T\boldsymbol{x}_i + w_0) - 1] = 0, \qquad i = 1, 2, 3, 4$$

Since, in the general case where $u < \frac{1}{2}$, all λ_i's are positive, we consider the equation

$$[y_i(\boldsymbol{w}^T\boldsymbol{x}_i + w_0) - 1] = 0$$

for all the vectors. Specifically, it is

$$1(-w_1 + w_2 + w_0) - 1 = 0, \quad for\ \boldsymbol{x}_1$$

$$1(-w_1 - w_2 + w_0) - 1 = 0, \quad for\ \boldsymbol{x}_2$$

$$(-1)(w_1 - w_2 + w_0) - 1 = 0, \quad for\ \boldsymbol{x}_3$$

$$(-1)(w_1 + w_2 + w_0) - 1 = 0, \quad for\ \boldsymbol{x}_4$$

where $\boldsymbol{w} = [w_1,\ w_2]^T = [-1,\ 0]^T$. Substituting the values of w_1 and w_2 in the above equations, we finally obtain that $w_0 = 0$, and,thus, the solution hyperplane is now completely specified.