The second more successful approach is to add the term $(1/2)w^\top w$ to the objective function and to drop the constraint $w^\top w \leq 1$. There are several variants of this method, depending on the choice of the regularizing term involving ϵ and ξ (linear or quadratic), how the margin is dealt with (implicitly with the term 1 or explicitly with a term η), and whether the term $(1/2)b^2$ is added to the objective function or not.

These methods all share the property that if the primal problem has an optimal solution with $w \neq 0$, then the dual problem always determines w, and then under mild conditions which we call standard margin hypotheses, b and η can be determined. Then ϵ and ξ can be determined using the constraints that are active. When $(1/2)b^2$ is added to the objective function, b is determined by the equation

$$b = -(\mathbf{1}_p^\top \lambda - \mathbf{1}_q^\top \mu).$$

All these problems are convex and the constraints are qualified, so the duality gap is zero, and if the primal has an optimal solution with $w \neq 0$, then it follows that $\eta \geq 0$.

We now consider five variants in more details.

(1) **Basic soft margin SVM**: (SVM$_{s2}$).

This is the optimization problem in which the regularization term $K \begin{pmatrix} \epsilon^\top & \xi^\top \end{pmatrix} \mathbf{1}_{p+q}$ is linear and the margin δ is given by $\delta = 1/\left\|w\right\|$:

$$\begin{aligned}
&\text{minimize} \quad \frac{1}{2} w^\top w + K \begin{pmatrix} \epsilon^\top & \xi^\top \end{pmatrix} \mathbf{1}_{p+q} \\
&\text{subject to} \\
&\qquad w^\top u_i - b \geq 1 - \epsilon_i, \quad \epsilon_i \geq 0 \qquad i = 1, \ldots, p \\
&\qquad - w^\top v_j + b \geq 1 - \xi_j, \quad \xi_j \geq 0 \qquad j = 1, \ldots, q.
\end{aligned}$$

This problem is the classical one discussed in all books on machine learning or pattern analysis, for instance Vapnik [182], Bishop [23], and Shawe–Taylor and Christianini [159]. It is shown in Section 54.3 that the dual program is

Dual of the Basic soft margin SVM: (SVM$_{s2}$):

$$\begin{aligned}
&\text{minimize} \quad \frac{1}{2} \begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} X^\top X \begin{pmatrix} \lambda \\ \mu \end{pmatrix} - \begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} \mathbf{1}_{p+q} \\
&\text{subject to} \\
&\qquad \sum_{i=1}^p \lambda_i - \sum_{j=1}^q \mu_j = 0 \\
&\qquad 0 \leq \lambda_i \leq K, \quad i = 1, \ldots, p \\
&\qquad 0 \leq \mu_j \leq K, \quad j = 1, \ldots, q.
\end{aligned}$$