This is the classical problem discussed in all books on machine learning or pattern analysis, for instance Vapnik [182], Bishop [23], and Shawe–Taylor and Christianini [159]. The trivial solution where all variables are 0 is ruled out because of the presence of the 1 in the inequalities, but it is not clear that if (w, b, ϵ, ξ) is an optimal solution, then $w \neq 0$.

We prove that if the primal problem has an optimal solution (w, ϵ, ξ, b) with $w \neq 0$, then w is determined by any optimal solution (λ, μ) of the dual. We also prove that there is some i for which $\lambda_i > 0$ and some j for which $\mu_j > 0$. Under a mild hypothesis that we call the **Standard Margin Hypothesis**, b can be found.

Note that this framework is still somewhat sensitive to outliers because the penalty for misclassification is linear in ϵ and ξ.

First we write the constraints in matrix form. The $2(p+q) \times (n+p+q+1)$ matrix C is written in block form as

$$C = \begin{pmatrix} X^\top & -I_{p+q} & \begin{matrix} \mathbf{1}_p \\ -\mathbf{1}_q \end{matrix} \\ 0_{p+q,n} & -I_{p+q} & 0_{p+q} \end{pmatrix},$$

where X is the $n \times (p+q)$ matrix

$$X = \begin{pmatrix} -u_1 & \cdots & -u_p & v_1 & \cdots & v_q \end{pmatrix},$$

and the constraints are expressed by

$$\begin{pmatrix} X^\top & -I_{p+q} & \begin{matrix} \mathbf{1}_p \\ -\mathbf{1}_q \end{matrix} \\ 0_{p+q,n} & -I_{p+q} & 0_{p+q} \end{pmatrix} \begin{pmatrix} w \\ \epsilon \\ \xi \\ b \end{pmatrix} \leq \begin{pmatrix} -\mathbf{1}_{p+q} \\ 0_{p+q} \end{pmatrix}.$$

The objective function $J(w, \epsilon, \xi, b)$ is given by

$$J(w, \epsilon, \xi, b) = \frac{1}{2} w^\top w + K \begin{pmatrix} \epsilon^\top & \xi^\top \end{pmatrix} \mathbf{1}_{p+q}.$$

The Lagrangian $L(w, \epsilon, \xi, b, \lambda, \mu, \alpha, \beta)$ with $\lambda, \alpha \in \mathbb{R}^p_+$ and with $\mu, \beta \in \mathbb{R}^q_+$ is given by

$$\begin{aligned} L(w, \epsilon, \xi, b, \lambda, \mu, \alpha, \beta) &= \frac{1}{2} w^\top w + K \begin{pmatrix} \epsilon^\top & \xi^\top \end{pmatrix} \mathbf{1}_{p+q} \\ &\quad + \begin{pmatrix} w^\top & \begin{pmatrix} \epsilon^\top & \xi^\top \end{pmatrix} & b \end{pmatrix} C^\top \begin{pmatrix} \lambda \\ \mu \\ \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \mathbf{1}_{p+q}^\top & 0_{p+q}^\top \end{pmatrix} \begin{pmatrix} \lambda \\ \mu \\ \alpha \\ \beta \end{pmatrix}. \end{aligned}$$