Then it is easy to see that we can compute b and δ using the following averaging formulae:

$$b = w^\top \left(\Big(\sum_{i \in I_\lambda} u_i \Big) / |I_\lambda| + \Big(\sum_{j \in I_\mu} v_j \Big) / |I_\mu| \right) /2$$

$$\delta = w^\top \left(\Big(\sum_{i \in I_\lambda} u_i \Big) / |I_\lambda| - \Big(\sum_{j \in I_\mu} v_j \Big) / |I_\mu| \right) /2.$$

As we said earlier, the hypotheses of Theorem 50.17(2) hold, so *if* the primal problem (SVM$_{s1}$) has an optimal solution with $w \neq 0$, *then* the dual problem has a solution too, and the duality gap is zero. Therefore, for optimal solutions we have

$$L(w, \epsilon, \xi, b, \delta, \lambda, \mu, \alpha, \beta, \gamma) = G(\lambda, \mu, \alpha, \beta, \gamma),$$

which means that

$$-\delta + K \bigg(\sum_{i=1}^p \epsilon_i + \sum_{j=1}^q \xi_j \bigg) = - \left(\begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} X^\top X \begin{pmatrix} \lambda \\ \mu \end{pmatrix} \right)^{1/2},$$

so we get

$$\delta = K \bigg(\sum_{i=1}^p \epsilon_i + \sum_{j=1}^q \xi_j \bigg) + \left(\begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} X^\top X \begin{pmatrix} \lambda \\ \mu \end{pmatrix} \right)^{1/2}.$$

Therefore, we confirm that $\delta \geq 0$.

It is important to note that the objective function of the dual program

$$-G(\lambda, \mu) = \left(\begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} X^\top X \begin{pmatrix} \lambda \\ \mu \end{pmatrix} \right)^{1/2}$$

only involves the inner products of the u_i and the v_j through the matrix $X^\top X$, and similarly, the equation of the optimal hyperplane can be written as

$$\sum_{i=1}^p \lambda_i u_i^\top x - \sum_{j=1}^q \mu_j v_j^\top x - \left(\begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} X^\top X \begin{pmatrix} \lambda \\ \mu \end{pmatrix} \right)^{1/2} b = 0,$$

an expression that only involves inner products of x with the u_i and the v_j and inner products of the u_i and the v_j.

As explained at the beginning of this chapter, this is a key fact that allows a generalization of the support vector machine using the method of *kernels*. We can define the following "kernelized" version of Problem (SVM$_{s1}$):