should hold.

The goal is to design an objective function that minimizes ϵ and ξ and maximizes δ. The optimization problem should also solve for w and b, and for this some constraint has to be placed on w. Another goal is to try to use the dual program to solve the optimization problem, because the solutions involve inner products, and thus the problem is amenable to a generalization using kernel functions.

The first attempt, which is to use the objective function

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

and the constraint $w^\top w \leq 1$, does not work very well because this constraint needs to be guarded by a Lagrange multiplier $\gamma \geq 0$, and as a result, minimizing the Lagrangian L to find the dual function G gives an equation for solving w of the form

$$2\gamma w = -X^\top \begin{pmatrix} \lambda \\ \mu \end{pmatrix},$$

but if the sets $\{u_i\}_{i=1}^p$ and $\{v_j\}_{j=1}^q$ are not linearly separable, then an optimal solution may occurs for $\gamma = 0$, in which case it is impossible to determine w. This is Problem (SVM_{s1}) considered in Section 54.1.

Soft margin SVM (SVM_{s1}):

$$\begin{aligned}
&\text{minimize} \quad -\delta + K\left(\sum_{i=1}^p \epsilon_i + \sum_{j=1}^q \xi_j\right) \\
&\text{subject to} \\
&\qquad w^\top u_i - b \geq \delta - \epsilon_i, \quad \epsilon_i \geq 0 \qquad i = 1, \ldots, p \\
&\qquad -w^\top v_j + b \geq \delta - \xi_j, \quad \xi_j \geq 0 \qquad j = 1, \ldots, q \\
&\qquad w^\top w \leq 1.
\end{aligned}$$

It is customary to write $\ell = p + q$.

It is shown in Section 54.1 that the dual program is equivalent to the following minimization program: