54.13 Soft Margin SVM; (SVM_{s4})

In this section we consider the version of Problem $(\mathrm{SVM}_{s2'})$ in which instead of using the function $K\left(\sum_{i=1}^p \epsilon_i + \sum_{j=1}^q \xi_j\right)$ as a regularizing function we use the quadratic function $K(\|\epsilon\|_2^2 + \|\xi\|_2^2)$.

Soft margin SVM (SVM_{s4}):

$$\begin{aligned}
&\text{minimize} \quad \frac{1}{2}w^\top w + (p+q)K_s\left(-\nu\eta + \frac{1}{p+q}(\epsilon^\top\epsilon + \xi^\top\xi)\right)\\
&\text{subject to}\\
&\qquad w^\top u_i - b \geq \eta - \epsilon_i, \qquad i = 1, \ldots, p\\
&\qquad -w^\top v_j + b \geq \eta - \xi_j, \qquad j = 1, \ldots, q\\
&\qquad \eta \geq 0,
\end{aligned}$$

where ν and K_s are two given positive constants. As we saw earlier, theoretically, it is convenient to pick $K_s = 1/(p+q)$. When writing a computer program, it is preferable to assume that K_s is arbitrary. In this case ν needs to be replaced by $(p+q)K_s\nu$ in all the formulae obtained with $K_s = 1/(p+q)$.

The new twist with this formulation of the problem is that if $\epsilon_i < 0$, then the corresponding inequality $w^\top u_i - b \geq \eta - \epsilon_i$ implies the inequality $w^\top u_i - b \geq \eta$ obtained by setting ϵ_i to zero while reducing the value of $\|\epsilon\|^2$, and similarly if $\xi_j < 0$, then the corresponding inequality $-w^\top v_j + b \geq \eta - \xi_j$ implies the inequality $-w^\top v_j + b \geq \eta$ obtained by setting ξ_j to zero while reducing the value of $\|\xi\|^2$. Therefore, if (w, b, ϵ, ξ) is an optimal solution of Problem (SVM_{s4}), it is not necessary to restrict the slack variables ϵ_i and ξ_j to the nonnegative, which simplifies matters a bit. In fact, we will see that for an optimal solution, $\epsilon = \lambda/(2K_s)$ and $\xi = \mu/(2K_s)$. The variable η can also be determined by expressing that the duality gap is zero.

One of the advantages of this methods is that ϵ is determined by λ, ξ is determined by μ, and η and b are determined by λ and μ. This method *does not* require support vectors to compute b. We can omit the constraint $\eta \geq 0$, because for an optimal solution it can be shown using duality that $\eta \geq 0$; see Section 54.14.

A drawback of Program (SVM_{s4}) is that for fixed K_s, the quantity $\delta = \eta/\|w\|$ and the hyperplanes $H_{w,b}, H_{w,b+\eta}$ and $H_{w,b-\eta}$ are independent of ν. This will be shown in Theorem 54.8. Thus this method is less flexible than $(\mathrm{SVM}_{s2'})$ and (SVM_{s3}).