If the sets of points $\{u_1, \ldots, u_p\}$ and $\{v_1, \ldots, v_q\}$ are not linearly separable (with $u_i, v_j \in \mathbb{R}^n$), we can use a trick from linear programming which is to introduce nonnegative "slack variables" $\epsilon = (\epsilon_1, \ldots, \epsilon_p) \in \mathbb{R}^p$ and $\xi = (\xi_1, \ldots, \xi_q) \in \mathbb{R}^q$ to relax the "hard" constraints

$$\begin{aligned} w^\top u_i - b &\geq \delta & \quad i &= 1, \ldots, p \\ -w^\top v_j + b &\geq \delta & \quad j &= 1, \ldots, q \end{aligned}$$

of Problem (SVM$_{h1}$) from Section 50.5 to the "soft" constraints

$$\begin{aligned} w^\top u_i - b &\geq \delta - \epsilon_i, \quad \epsilon_i \geq 0 \quad i = 1, \ldots, p \\ -w^\top v_j + b &\geq \delta - \xi_j, \quad \xi_j \geq 0 \quad j = 1, \ldots, q. \end{aligned}$$

Recall that $w \in \mathbb{R}^n$ and $b, \delta \in \mathbb{R}$.

If $\epsilon_i > 0$, the point u_i may be misclassified, in the sense that it can belong to the margin (the slab), or even to the wrong half-space classifying the negative (red) points. See Figures 54.5 (2) and (3). Similarly, if $\xi_j > 0$, the point v_j may be misclassified, in the sense that it can belong to the margin (the slab), or even to the wrong half-space classifying the positive (blue) points. We can think of ϵ_i as a measure of how much the constraint $w^\top u_i - b \geq \delta$ is violated, and similarly of ξ_j as a measure of how much the constraint $-w^\top v_j + b \geq \delta$ is violated. If $\epsilon = 0$ and $\xi = 0$, then we recover the original constraints. By making ϵ and ξ large enough, these constraints can always be satisfied. We add the constraint $w^\top w \leq 1$ and we minimize $-\delta$.

If instead of the constraints of Problem (SVM$_{h1}$) we use the hard constraints

$$\begin{aligned} w^\top u_i - b &\geq 1 & \quad i &= 1, \ldots, p \\ -w^\top v_j + b &\geq 1 & \quad j &= 1, \ldots, q \end{aligned}$$

of Problem (SVM$_{h2}$) (see Example 50.6), then we relax to the soft constraints

$$\begin{aligned} w^\top u_i - b &\geq 1 - \epsilon_i, \quad \epsilon_i \geq 0 \quad i = 1, \ldots, p \\ -w^\top v_j + b &\geq 1 - \xi_j, \quad \xi_j \geq 0 \quad j = 1, \ldots, q. \end{aligned}$$

In this case there is no constraint on w, but we minimize $(1/2)w^\top w$.

Ideally we would like to find a separating hyperplane that *minimizes the number of misclassified points*, which means that the variables ϵ_i and ξ_j should be as small as possible, but there is a trade-off in maximizing the margin (the thickness of the slab), and minimizing the number of misclassified points. This is reflected in the choice of the objective function, and there are several options, depending on whether we minimize a linear function of the variables ϵ_i and ξ_j, or a quadratic functions of these variables, or whether we include the term $(1/2)b^2$ in the objective function. These methods are known as *support vector classification* algorithms (for short *SVC* algorithms).