Note that if ν is chosen so that $\nu < 1/(p+q)$, then $p_f = q_f = 0$, which means that none of the data points are misclassified; in other words, the u_is and v_js are linearly separable. Thus we see that if the u_is and v_js are not linearly separable we must pick ν such that $1/(p+q) \leq \nu \leq 1$ for the method to succeed. In fact, by Proposition 54.5, we must choose ν so that

$$\frac{p_f + q_f}{p+q} \leq \nu \leq \frac{p_m + q_m}{p+q}.$$

Furthermore, in order to be able to determine b, we must have the strict inequality

$$\frac{p_f + q_f}{p+q} < \nu.$$

54.11 Existence of Support Vectors for (SVM_{s3})

The following proposition is the version of Proposition 54.2 for Problem (SVM_{s3}).

Proposition 54.6. *For every optimal solution $(w, b, \eta, \epsilon, \xi)$ of Problem (SVM_{s3}) with $w \neq 0$ and $\eta > 0$, if $\nu < 1$ and if no u_i is a support vector and no v_j is a support vector, then there is another optimal solution such that some u_{i_0} or some v_{j_0} is a support vector.*

Proof. We may assume that $K_s = 1/(p+q)$ and we proceed by contradiction. Thus we assume that for all $i \in \{1, \ldots, p\}$, if $\epsilon_i = 0$, then the constraint $w^\top u_i - b \geq \eta$ is not active, namely $w^\top u_i - b > \eta$, *and* for all $j \in \{1, \ldots, q\}$, if $\xi_j = 0$, then the constraint $-w^\top v_j + b \geq \eta$ is not active, namely $-w^\top v_j + b > \eta$.

Let $E_\lambda = \{i \in \{1, \ldots, p\} \mid \epsilon_i > 0\}$ and let $E_\mu = \{j \in \{1, \ldots, q\} \mid \xi_j > 0\}$. By definition, $p_{sf} = |E_\lambda|$, $q_{sf} = |E_\mu|$, $p_{sf} \leq p_f$ and $q_{sf} \leq q_f$, so by Proposition 54.1,

$$\frac{p_{sf} + q_{sf}}{p+q} \leq \frac{p_f + q_f}{p+q} \leq \nu.$$

Therefore, if $\nu < 1$, then $p_{sf} + q_{sf} < p + q$, which implies that either there is some $i \notin E_\lambda$ such that $\epsilon_i = 0$ or there is some $j \notin E_\mu$ such that $\xi_j = 0$.

By complementary slackness all the constraints for which $i \in E_\lambda$ and $j \in E_\mu$ are active, so our hypotheses are

$$\begin{aligned}
w^\top u_i - b &= \eta - \epsilon_i & \epsilon_i &> 0 & i &\in E_\lambda \\
-w^\top v_j + b &= \eta - \xi_j & \xi_j &> 0 & j &\in E_\mu \\
w^\top u_i - b &> \eta & & & i &\notin E_\lambda \\
-w^\top v_j + b &> \eta & & & j &\notin E_\mu,
\end{aligned}$$

and either there is some $i \notin E_\lambda$ or there is some $j \notin E_\mu$. Our strategy, as illustrated in Figures 54.8 and 54.9, is to increase the width η of the slab keeping the separating hyperplane unchanged. Let us pick θ such that

$$\theta = \min\{w^\top u_i - b - \eta,\ -w^\top v_j + b - \eta \mid i \notin E_\lambda, j \notin E_\mu\}.$$