which yields

$$p_m \geq \frac{\nu m}{2}.$$

A similar reasoning applies if $\mu_i > 0$.

(3) This follows immediately from (1). $\square$

Proposition 56.2 yields bounds on ν, namely

$$\max\left\{\frac{2p_f}{m}, \frac{2q_f}{m}\right\} \leq \nu \leq \min\left\{\frac{2p_m}{m}, \frac{2q_m}{m}\right\},$$

with $p_f \leq p_m$, $q_f \leq q_m$, $p_f + q_f \leq m$ and $p_m + q_m \leq m$. Also, $p_f = q_f = 0$ means that the ϵ-slab is wide enough so that there are no errors (no points strictly outside the slab).

Observe that a small value of ν keeps p_f and q_f small, which is achieved if the ϵ-slab is wide. A large value of ν allows p_m and q_m to be fairly large, which is achieved if the ϵ-slab is narrow. Thus the smaller ν is, the wider the ϵ-slab is, and the larger ν is, the narrower the ϵ-slab is.

56.2 Existence of Support Vectors

We now consider the issue of the existence of support vectors. We will show that in the generic case, for any optimal solution for which $\epsilon > 0$, there is some support vector on the blue margin *and* some support vector on the red margin. Here generic means that there is an optimal solution for some $\nu < (m-1)/m$.

If the data set (X, y) is well fit by some affine function $f(x) = w^\top x + b$, in the sense that for many pairs (x_i, y_i) we have $y_i = w^\top x_i + b$ and the ℓ^1-error

$$\sum_{i=1}^{m} |w^\top x_i + b - y_i|$$

is small, then an optimal solution may have $\epsilon = 0$. Geometrically, many points (x_i, y_i) belong to the hyperplane $H_{w,b}$. The situation in which $\epsilon = 0$ corresponds to minimizing the ℓ^1-error with a quadratic penalization of w. This is a sort of dual of lasso. The fact that the affine function $f(x) = w^\top x + b$ fits perfectly many points corresponds to the fact that an ℓ^1-minimization tends to encourage sparsity. In this case, if C is chosen too small, it is possible that all points are errors (although "small") and there are no support vectors. But if C is large enough, the solution will be sparse and there will be *many* support vectors on the hyperplane $H_{w,b}$.

Let $E_\lambda = \{i \in \{1, \ldots, m\} \mid \xi_i > 0\}$, $E_\mu = \{j \in \{1, \ldots, m\} \mid \xi'_j > 0\}$, $p_{sf} = |E_\lambda|$ and $q_{sf} = |E_\mu|$. Obviously, E_λ and E_μ are disjoint.

Given any real numbers u, v, x, y, if $\max\{u, v\} < \min\{x, y\}$, then $u < x$ and $v < y$. This is because $u, v \leq \max\{u, v\} < \min\{x, y\} \leq x, y$.