Observe that if $\nu > 1$, then an optimal solution of the above program must yield $\epsilon = 0$. Indeed, if $\epsilon > 0$, we can reduce it by a small amount $\delta > 0$ and increase $\xi_i + \xi_i'$ by δ to still satisfy the constraints, but the objective function changes by the amount $-\nu\delta + \delta$, which is negative since $\nu > 1$, so $\epsilon > 0$ is not optimal.

Driving ϵ to zero is not the intended goal, because typically the data is not noise free so very few pairs (x_i, y_i) will satisfy the equation $w^\top x_i + b = y_i$, and then many pairs (x_i, y_i) will correspond to an error ($\xi_i > 0$ or $\xi_i' > 0$). Thus, *typically we assume that* $0 < \nu \leq 1$.

To construct the Lagrangian, we assign Lagrange multipliers $\lambda_i \geq 0$ to the constraints $w^\top x_i + b - y_i \leq \epsilon + \xi_i$, Lagrange multipliers $\mu_i \geq 0$ to the constraints $-w^\top x_i - b + y_i \leq \epsilon + \xi_i'$, Lagrange multipliers $\alpha_i \geq 0$ to the constraints $\xi_i \geq 0$, Lagrange multipliers $\beta_i \geq 0$ to the constraints $\xi_i' \geq 0$, and the Lagrange multiplier $\gamma \geq 0$ to the constraint $\epsilon \geq 0$. The Lagrangian is

$$\begin{aligned}L(w, b, \lambda, \mu, \gamma, \xi, \xi', \epsilon, \alpha, \beta) &= \frac{1}{2} w^\top w + C\Bigg(\nu\epsilon + \frac{1}{m}\sum_{i=1}^m (\xi_i + \xi_i')\Bigg) - \gamma\epsilon - \sum_{i=1}^m (\alpha_i \xi_i + \beta_i \xi_i') \\ &\quad + \sum_{i=1}^m \lambda_i (w^\top x_i + b - y_i - \epsilon - \xi_i) + \sum_{i=1}^m \mu_i (-w^\top x_i - b + y_i - \epsilon - \xi_i').\end{aligned}$$

The Lagrangian can also be written as

$$\begin{aligned}L(w, b, \lambda, \mu, \gamma, \xi, \xi', \epsilon, \alpha, \beta) &= \frac{1}{2} w^\top w + w^\top \left(\sum_{i=1}^m (\lambda_i - \mu_i) x_i\right) + \epsilon\left(C\nu - \gamma - \sum_{i=1}^m (\lambda_i + \mu_i)\right) \\ &+ \sum_{i=1}^m \xi_i \left(\frac{C}{m} - \lambda_i - \alpha_i\right) + \sum_{i=1}^m \xi_i' \left(\frac{C}{m} - \mu_i - \beta_i\right) + b\left(\sum_{i=1}^m (\lambda_i - \mu_i)\right) - \sum_{i=1}^m (\lambda_i - \mu_i) y_i.\end{aligned}$$

To find the dual function $G(\lambda, \mu, \gamma, \alpha, \beta)$, we minimize $L(w, b, \lambda, \mu, \gamma, \xi, \xi', \epsilon, \alpha, \beta)$ with respect to the primal variables w, ϵ, b, ξ and ξ'. Observe that the Lagrangian is convex, and since $(w, \epsilon, \xi, \xi', b) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}$, a convex open set, by Theorem 40.13, the Lagrangian has a minimum iff $\nabla L_{w,\epsilon,b,\xi,\xi'} = 0$, so we compute the gradient $\nabla L_{w,\epsilon,b,\xi,\xi'}$. We obtain

$$\nabla L_{w,\epsilon,b,\xi,\xi'} = \begin{pmatrix} w + \sum_{i=1}^m (\lambda_i - \mu_i) x_i \\ C\nu - \gamma - \sum_{i=1}^m (\lambda_i + \mu_i) \\ \sum_{i=1}^m (\lambda_i - \mu_i) \\ \frac{C}{m} - \lambda - \alpha \\ \frac{C}{m} - \mu - \beta \end{pmatrix},$$

where

$$\left(\frac{C}{m} - \lambda - \alpha\right)_i = \frac{C}{m} - \lambda_i - \alpha_i, \quad \text{and} \quad \left(\frac{C}{m} - \mu - \beta\right)_i = \frac{C}{m} - \mu_i - \beta_i.$$