Then it is easy to see that we can compute ϵ using the following averaging formulae: if $I_\lambda \neq \emptyset$, then

$$\epsilon = w^\top \Big(\sum_{i\in I_\lambda} x_i\Big)/|I_\lambda| + b - \Big(\sum_{i\in I_\lambda} y_i\Big)/|I_\lambda|,$$

and if $I_\mu \neq \emptyset$, then

$$\epsilon = -w^\top \Big(\sum_{j\in I_\mu} x_j\Big)/|I_\mu| - b + \Big(\sum_{i\in I_\mu} y_i\Big)/|I_\mu|.$$

The second method uses duality. Under a mild condition, expressing that the duality gap is zero, we can determine ϵ in terms of λ, μ and b. This is because points x_i that fail the margin, which means that $\lambda_i = C/m$ or $\mu_i = C/m$, are the only points for which $\xi_i > 0$ or $\xi_i' > 0$. But in this case we have an active constraint

$$w^\top x_i + b - y_i = \epsilon + \xi_i \qquad (*_\xi)$$

or

$$-w^\top x_i - b + y_i = \epsilon + \xi_i', \qquad (*_{\xi'})$$

so ξ_i and ξ_i' can be expressed in terms of w and b. Since the duality gap is zero for an optimal solution, the optimal value of the primal is equal to the optimal value of the dual. Using the fact that

$$\begin{aligned} w &= X^\top(\mu - \lambda) \\ b &= -(\mathbf{1}_m^\top\lambda - \mathbf{1}_m^\top\mu) = \begin{pmatrix}\lambda^\top & \mu^\top\end{pmatrix}\begin{pmatrix}-\mathbf{1}_m \\ \mathbf{1}_m\end{pmatrix}\end{aligned}$$

we obtain an expression for the optimal value of the primal. First we have

$$\begin{aligned} \frac{1}{2}w^\top w + \frac{1}{2}b^2 &= \frac{1}{2}(\lambda^\top - \mu^\top)XX^\top(\lambda - \mu) + \frac{1}{2}\begin{pmatrix}\lambda^\top & \mu^\top\end{pmatrix}\begin{pmatrix}\mathbf{1}_m\mathbf{1}_m^\top & -\mathbf{1}_m\mathbf{1}_m^\top \\ -\mathbf{1}_m\mathbf{1}_m^\top & \mathbf{1}_m\mathbf{1}_m^\top\end{pmatrix}\begin{pmatrix}\lambda \\ \mu\end{pmatrix} \\ &= \frac{1}{2}\begin{pmatrix}\lambda^\top & \mu^\top\end{pmatrix}\left(P + \begin{pmatrix}\mathbf{1}_m\mathbf{1}_m^\top & -\mathbf{1}_m\mathbf{1}_m^\top \\ -\mathbf{1}_m\mathbf{1}_m^\top & \mathbf{1}_m\mathbf{1}_m^\top\end{pmatrix}\right)\begin{pmatrix}\lambda \\ \mu\end{pmatrix}, \end{aligned}$$

with

$$P = \begin{pmatrix} XX^\top & -XX^\top \\ -XX^\top & XX^\top \end{pmatrix}.$$

Let K_λ and K_μ be the sets of indices corresponding to points failing the margin,

$$\begin{aligned} K_\lambda &= \{i \in \{1, \ldots, m\} \mid \lambda_i = C/m\} \\ K_\mu &= \{i \in \{1, \ldots, m\} \mid \mu_i = C/m\}. \end{aligned}$$