Elementary SMO Optimization Step

Thus, consider optimizing over two variables α_i and α_j with all other variables fixed. Using the shorthand $K_{ij} := k(x_i, x_j)$, (8.13)–(8.15)) then reduce to (up to a constant)

$$\begin{aligned} \underset{\alpha_i,\alpha_j}{\text{minimize}} \quad & \frac{1}{2}\left[\alpha_i^2 K_{ii} + \alpha_j^2 K_{jj} + 2\alpha_i\alpha_j K_{ij}\right] + c_i\alpha_i + c_j\alpha_j \\ \text{subject to} \quad & \alpha_i + \alpha_j = \gamma \\ & 0 \le \alpha_i, \alpha_j \le \tfrac{1}{\nu m} \end{aligned} \tag{8.22}$$

in analogy to (10.63) below. Here the constants c_i, c_j, and γ are defined as follows;

$$c_i := \sum_{l\neq i,j}^{m} \alpha_l K_{il},\; c_j := \sum_{l\neq i,j}^{m} \alpha_l K_{jl},\; \text{and } \gamma = 1 - \sum_{l\neq i,j}^{m} \alpha_l. \tag{8.23}$$

To find the minimum, we use $\alpha_i + \alpha_j = \gamma$. This allows us to obtain a constrained optimization problem in α_i alone by elimination of α_j. For convenience we introduce $\chi := K_{ii} + K_{jj} - 2K_{ij}$.

$$\begin{aligned} \underset{\alpha_i}{\text{minimize}} \quad & \alpha_i^2\chi + 2\alpha_i\left(c_i - c_j + \gamma(K_{ij} - K_{jj})\right) \\ \text{subject to} \quad & L \le \alpha_i \le H, \text{ where } L = \max(0, \gamma - 1/(\nu m)) \text{ and } H = \min(1/(\nu m), \gamma). \end{aligned}$$

Without going into details (a similar calculation can be found in Section 10.5.1) the minimizer α_i of this optimization problem is given by

$$\alpha_i = \min(\max(L, \tilde{\alpha}_i), H). \tag{8.24}$$

where $\tilde{\alpha}_i$, the unconstrained solution, is given by

$$\begin{aligned} \tilde{\alpha}_i &= \alpha_i^{\text{old}} + \chi^{-1}\left(c_j - c_i + K_{jj}\alpha_j^{\text{old}} + K_{ij}\left(\alpha_j^{\text{old}} - \alpha_i^{\text{old}}\right) - K_{ii}\alpha_i^{\text{old}}\right) & (8.25) \\ &= \alpha_i^{\text{old}} + \chi^{-1}\left(f^{\text{old}}(x_j) - f^{\text{old}}(x_i)\right). & (8.26) \end{aligned}$$

Finally, α_j can be obtained via $\alpha_j = \gamma - \alpha_i$. Eq. (8.26) tells us that the change in α_i will depend on the difference between the values $f(x_i)$ and $f(x_j)$. The less close these values are, i.e., the larger the difference in the distances to the hyperplane, the larger the possible change in the set of variables. Note, however, that there is no guarantee that the actual change in α_i will indeed be large, since α_i has to satisfy the constraint $L \le \alpha_i \le H$. Finally, the size of χ plays an important role, too (for the case of $\chi = 0$ see Lemma 10.3). The larger it is, the smaller the likely change in α_i.

We next briefly describe how to do the overall optimization.

Initialization of the Algorithm We start by setting a random fraction ν of all α_i to $1/(\nu m)$. If νm is not an integer, then one of the examples is set to a value in $(0, 1/(\nu m))$ to ensure that $\sum_i \alpha_i = 1$. Furthermore, we set the initial ρ to $\max\{f(x_i)|i \in [m],\ \alpha_i > 0\}$.

Optimization Algorithm We then select the first variable for the elementary optimization step in one of two following ways. Here, we use the shorthand SV_{nb} for