can only be achieved by defining new kernels as linear combinations of differently scaled kernels. This is due to the fact that once a regularization operator is chosen, the solution minimizing the regularized risk function has to expanded into the corresponding Green's functions of P^*P (Chapter 4). In these cases, a possible way out is to resort to the LP version (Section 9.4). A final area of research left out of this chapter is the problem of estimating the values of functions at given test points, sometimes referred to as transduction [103].

9.8 Problems

9.1 (Product of SVR Lagrange Multipliers [561] •) *Show that for $\varepsilon > 0$, the solution of the SVR dual problem satisfies*

$$\alpha_i \alpha_i^* = 0 \tag{9.58}$$

for all $i = 1, \ldots, m$. Prove it either directly from (9.17), or from the KKT conditions.

Show that for $\varepsilon = 0$, we can always find a solution which satisfies (9.58) and which is optimal, by subtracting $\min\{\alpha_i, \alpha_i^*\}$ from both multipliers.

Give a mechanical interpretation of this result, in terms of forces on the SVs (cf. Chapter 7).

9.2 (SV Regression with Fewer Slack Variables ••) *Prove geometrically that in SV regression, we always have $\xi_i \xi_i^* = 0$. Argue that it is therefore sufficient to just introduce slacks ξ_i and use them in both (9.9) and (9.10). Derive the dual problem and show that it is identical to (9.17) except for a modified constraint $0 \leq \alpha_i + \alpha_i^* \leq C$. Using the result of Problem 9.1, prove that this problem is equivalent to (9.10).*

Hint: although the number of slacks is half of the original quantity, you still need both α_i and α_i^* to deal with the constraints.

9.3 (ν-Property from the Primal Objective Function •) *Try to understand the ν-property from the primal objective function (9.31). Assume that at the point of the solution, $\varepsilon > 0$, and set $(\partial/\partial\varepsilon)\tau(\mathbf{w}, \varepsilon)$ equal to 0.*

9.4 (One-Sided Regression ••) *Consider a situation where you are seeking a flat function that lies above all of the data points; that is, a regression that only measures errors in one direction. Formulate an SV algorithm by starting with the linear case and later introducing kernels. Generalize to the soft margin case, using the ν-trick. Discuss the applicability of such an algorithm. Also discuss how this algorithm is related to ν-SVR using different values of ν for the two sides of the tube.*

9.5 (Basis Pursuit ••) *Formulate a basis pursuit variant of SV regression, where, starting from zero, SVs are added iteratively in a greedy way (cf. [577]).*

9.6 (SV Regression with Hard Constraints •) *Derive dual programming problems for variants of ε-SVR and ν-SVR where all points are required to lie inside the ε-tubes (in*