Sparsity

6.18 Justify the following statement:

A support vector machine solver is sparse, but the Gram associated with the machine is rarely sparse.

6.19 The quadratic programming routine in a support vector machine solver provides the basis for splitting the training examples into three categories. Define these three categories, and use a two-dimensional figure to illustrate how the splitting is performed.

Metrics

6.20 With different algortihms being developed for accelerating support vector machine solvers, it is important that we formulate metrics for comparing the performance of these different algorithms. Develop a set of metrics that could be used for dealing with this practical issue.

Reproducing-kernel Hilbert space

6.21 Let $k(\mathbf{x}_i, \boldsymbol{.})$ and $k(\mathbf{x}_j, \boldsymbol{.})$ denote a pair of kernels, where $i, j = 1, 2, \ldots, N$. The vectors $\mathbf{x}_i$ and $\mathbf{x}_j$ have the same dimensionality. Show that

$$k(\mathbf{x}_i, \boldsymbol{.})k(\mathbf{x}_j, \boldsymbol{.}) = k(\mathbf{x}_i, \mathbf{x}_j)$$

where the expression on the left-hand side is an inner-product kernel.

6.22 Equations (6.77), (6.78), and (6.79) describe three important properties of the inner product $\langle f, g \rangle$, defined in Eq. (6.75). Prove the properties described in those three equations.

6.23 Justify the following statement:

If a reproducing kernel $k(\mathbf{x}, \mathbf{x}')$ exists, then that kernel is unique.

Computer experiments

6.24 This experiment investigates the scenario where the two moons in Fig. 1.8 overlap and are therefore nonseparable.

(a) Repeat the second part of the experiment in Fig. 6.7, for which the vertical separation between the two moons was fixed at $d = -6.5$. Experimentally, determine the value of parameter C for which the classification error rate is reduced to a minimum.

(b) Reduce the vertical separation between the two moons further by setting $d = -6.75$, for which the classification error rate is expected to be higher than that for $d = -6.5$. Experimentally, determine the value of parameter C for which the error rate is reduced to a minimum.

Comment on the results obtained for both parts of the experiment.

6.25 Among the supervised-learning algorithms studied thus far, the support vector machine stands out as the most powerful. In this problem, the performance of the support vector machine is to be challenged by using it to classify the two multicircular regions that constitute the "tightly fisted" structure shown in Fig. P6.24. The radii of the three concentric circles in this figure are $d_1 = 0.2$, $d_2 = 0.5$, and $d_3 = 0.8$.

(a) Generate 100 epochs, each of which consists of 200 randomly distributed training examples, and an equal number of test data for the two regions of Fig. P6.24.

(b) Train a support vector machine, assigning the value $C = 500$. Hence, construct the decision boundary computed by the machine.

(c) Test the network and thereby determine the classification error rate.

(d) Repeat the experiment for $C = 100$ and $C = 2{,}500$.

Comment on your results.

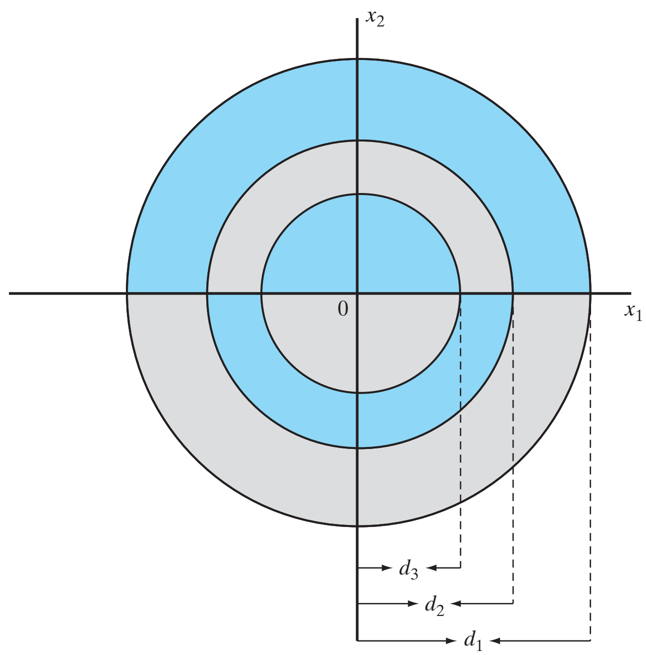

FIGURE P6.25