Lab 7

Learning in Reproducing Kernel Hilbert Spaces

Advanced Machine Learning

DATA 442/642

Exercise 1

For this exercise, the performance of the SVM is tested in the context of a two-class two-dimensional classification task. The data set comprises $N = 150$ points uniformly distributed in the region $[-5, 5] \times [-5, 5]$. For each point $\boldsymbol{x}_n = [x_{n,1}, x_{n,2}]^\top$, we compute

$$y_n = 0.05x_{n,1}^3 + 0.05x_{n,1}^2 + 0.05x_{n,1} + 0.05 + \eta,$$

where η stands for zero mean Gaussian noise of variance $\sigma_\eta^2 = 4$. The point is assigned to either of the two classes, depending on the value of the noise as well as its position with respect to the graph of the function

$$f(x) = 0.05x^3 + 0.05x^2 + 0.05x + 0.05$$

in the two-dimensional space. That is, if $x_{n,2} \geq y_n$, the point is assigned to class ω_1; otherwise, it is assigned to class ω_2.

(a) Plot the points $[x_{n,1}, x_{n,2}]$ using different colors for each class.

(b) Use SVM with the Gaussian kernel for $\sigma = 20$ and set $C = 1$. Plot the classifier and the margin. Moreover, find the support vectors (i.e., the points with nonzero Lagrange multipliers that contribute to the expansion of the classifier) and plot them as circled points.

(c) Repeat step (b) using $C = 0.5, 0.1, 0.05$.

(d) Repeat step (b) using $C = 5, 10, 50, 100$.

(e) Comment on the results.