Table 9.3 Results for the Boston housing benchmark; *top:* ν-SVR, *bottom:* ε-SVR. Abbreviation key: MSE: Mean squared errors, STD: standard deviation thereof (100 trials), Errors: fraction of training points outside the tube, SVs: fraction of training points which are SVs.

ν	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
automatic ε	2.6	1.7	1.2	0.8	0.6	0.3	0.0	0.0	0.0	0.0
MSE	9.4	8.7	9.3	9.5	10.0	10.6	11.3	11.3	11.3	11.3
STD	6.4	6.8	7.6	7.9	8.4	9.0	9.6	9.5	9.5	9.5
Errors	0.0	0.1	0.2	0.2	0.3	0.4	0.5	0.5	0.5	0.5
SVs	0.3	0.4	0.6	0.7	0.8	0.9	1.0	1.0	1.0	1.0

ε	0	1	2	3	4	5	6	7	8	9	10
MSE	11.3	9.5	8.8	9.7	11.2	13.1	15.6	18.2	22.1	27.0	34.3
STD	9.5	7.7	6.8	6.2	6.3	6.0	6.1	6.2	6.6	7.3	8.4
Errors	0.5	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SVs	1.0	0.6	0.4	0.3	0.2	0.1	0.1	0.1	0.1	0.1	0.1

9.6 Applications

Boston Housing Benchmark

Empirical studies using ε-SVR have shown excellent performance on the widely used Boston housing regression benchmark set [529]. Due to Proposition 9.3, the only difference between ν-SVR and standard ε-SVR lies in the fact that different parameters, ε vs. ν, have to be specified a priori. We now describe how the results obtained on this benchmark set change with the adjustment of parameters ε and ν. In our experiments, we kept all remaining parameters fixed, with C and the width $2s^2$ in $k(x, x') = \exp(-\|x - x'\|^2/(2s^2))$ chosen as in [480]: we used $2s^2 = 0.3 \cdot N$, where $N = 13$ is the input dimension, and $C/m = 10 \cdot 50$ (the original value of 10 was corrected since in the present case, the maximal y-value is 50 rather than 1). We performed 100 runs, where each time the overall set of 506 examples was randomly split into a training set of $m = 481$ examples and a test set of 25 examples (cf. [529]). Table 9.3 shows that over a wide range of ν (recall that only $0 \leq \nu \leq 1$ makes sense), we obtained performances which are close to the *best* performances that can be achieved using a value of ε selected a priori by looking at the test set.[6] Finally, note that although we did not use validation techniques to select the optimal values for C and $2s^2$, the performances are state of the art: Stitson et al. [529] report an MSE of 7.6 for ε-SVR using ANOVA kernels (cf. (13.13) in Section 13.6), and 11.7 for Bagging regression trees. Table 9.3 also shows that in this real-world application, ν can be used to control the fractions of SVs and errors.

Time Series Prediction

Time series prediction is a field that often uses regression techniques. The stan-

6. For a theoretical analysis of how to select the asymptotically optimal ν for a given noise model, cf. Section 3.4.4.