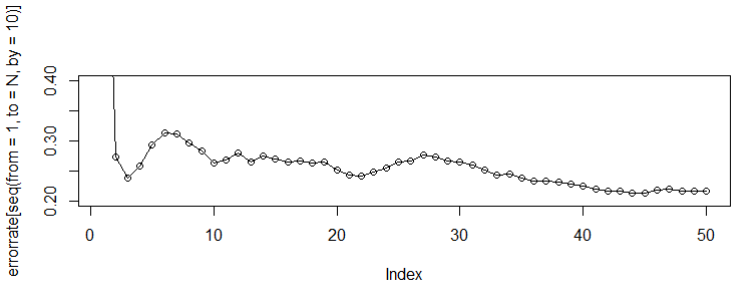
732A95 INTRODUCTION MACHINE LEARNING

FACIT 2018-01-11

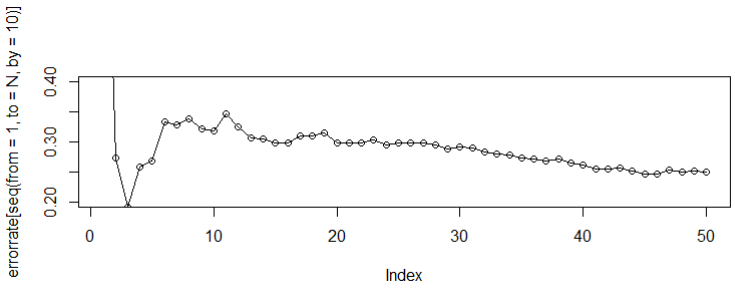
JOSE M. PEÑA
IDA, LINKÖPING UNIVERSITY, SWEDEN

ONLINE LEARNING

See the attached file for the code. For the setting (500, 0), the final error rate is 0.212, the SMV has 106 support vectors, and the error rate plot looks like the one below.



For the setting (500, -0.05), the final error rate is 0.248, the SMV has 46 support vectors, and the error rate plot looks like the one below.



The reason why the first setting gives better results is that every misclassified point is added as support vector, whereas in the second setting only seriously misclassified points are added, i.e. only those that are beta units beyond the decision boundary. This can also be seen by looking at the number of support vectors at the end of the two runs. Adding a misclassified point as support vector implies correcting the SVM so as to make that point (and other similar points) less likely to be misclassified in the future.

Finally, the setting (50,0) is the slowest because it is the one that removes more vectors. Removing is an expensive operation as one has to evaluate the result of removing each vector, and there are 50 of them.

NEURAL NETWORKS

See the attached file for the code. The best model is that with one layer of 10 neurons and threshold for early stopping equal to 4/1000. The generalization error is estimated by sampling additional test data, since we have access to the true function. Therefore, we conclude that more layers is not always better.