

Exercise

Support Vector Machines

Support vector machines: function `svm()` from the R package `e1071`

Take the bank data set (see previous exercises) for support vector machine classification. The goal is to predict if the client will subscribe a term deposit or not. This information is represented by the binary variable y (last one). Select randomly a training set of a reasonable size, compute the classifier, and evaluate the classifier based on the test set.

- (a) Apply `svm()` with the default parameters (radial basis kernel) to the training data. Predict the group membership for the test set and compute the misclassification rate.
- (b) Tune the parameters `gamma` and `cost` by using the function `tune.svm()`. Provide a grid of values for both parameters. Which parameters turn out as optimal? Plot the resulting object to visualize the misclassification error depending on the parameters.
- (c) Use the optimized parameters for `svm()`, build the model again for the training data, and evaluate for the test data. Did the error rate improve?
- (d) Try to improve the misclassification error of the “yes” clients (by keeping the overall misclassification error still small) with different strategies. You could either use balanced training data, or you could change the parameter `class.weights()` in the function. Note that also here you need to tune the parameters `gamma` and `cost`.