9.R Review Questions

In this problem, you will use simulation to evaluate (by Monte Carlo) the expected misclassification error rate given a particular generating model. Let y_i be equally divided between classes 0 and 1, and let $x_i \in \mathbb{R}^{10}$ be normally distributed.

Given $y_i = 0$, $x_i \sim N_{10}\left(0, I_{10}\right)$. Given $y_i = 1$, $x_i \sim N_{10}\left(\mu, I_{10}\right)$ with $\mu = (1, 1, 1, 1, 1, 0, 0, 0, 0, 0)$.

The N_{10} notation just means its a ten-dimensional Gaussian distribution; you can use the mvrnorm function in the MASS package to help generate the data. Now, we would like to know the expected test error rate if we fit an SVM to a sample of 50 random training points from class 1 and 50 more from class 0. We can calculate this to high precision by 1) generating a random training sample to train on, 2) evaluating the number of mistakes we make on a large test set, and then 3) repeating (1-2) many times and averaging the error rate for each trial.

Aside: in real life don't know the generating distribution, so we have to use resampling methods instead of the procedure described above.

For all of the following, please enter your error rate as a number between zero and 1 (e.g., 0.21 instead of 21 if the error rate is 21%).

9.R.1

1/1 point (graded)
Use svm in the e1071 package with the default settings (the default kernel is a radial kernel). What is the expected test error rate of this method (to within 10%)?

0.16350

✔ **Answer:** 0.16350

0.16350

Submit

9.R.2

1/1 point (graded)
Now fit an svm with a linear kernel (kernel = "linear"). What is the expected test error rate to within 10%?

| 0.15791 |

✔ **Answer:** 0.15791

0.15791

Explanation
An svm with a linear kernel does a little better here, since the best decision boundary is truly linear.

Submit

9.R.3

1/1 point (graded)
What is the expected test error for logistic regression? (to within 10%)

(Don't worry if you get errors saying the logistic regression did not converge.)

| 0.15750 |

✔ **Answer:** 0.15750

0.15750

Explanation
Logistic regression is similar to SVM with a linear kernel.

Submit

ⓘ   Answers are displayed within the problem