Machine Learning for Economics and Finance
Task 1: Logistic Regressions

02_Default_data

Ole Wilms

July 29, 2024

Task 1: Logistic Regressions

1.1 Randomly split the data into 7000 observations for training and 3000 observations for testing and set the seed to 1 before sampling the data. Call these two datasets *train_data* and *test_data* respectively. (Hint: use the code to split the data from 01 Auto_data_2.R or Auto_data_2.Rmd)

[ ]:

1.2 Fit a logistic regression of default on *income* using the *train_data*. Analyze the significance of the estimated coefficients.

[ ]:

1.3 Compute the *out-of-sample accuracy* and *error rate* and compare to the *in-sample statistics*. Do you think this is a good model to predict default?

[ ]:

1.4 Add balance as a predictor and compute the *out-of-sample error rate* and *accuracy*. Do you think this is a good model to predict *default*?

[ ]:

1.5 Compare the results for Task 1.4 to a model with only balance as a predictor. Which model would you choose?

[ ]:

1.6 Take the model from Task 1.4 but now re-estimate the model using different *seeds* to draw your *training* and *test data*. Does your *test error rate* change with the seed? What's going on here?

[ ]: