

Exercises

The following exercise offers you the opportunity to test your understanding of the materials presented to you in Lectures 5 and 7.

We use `PimaIndiansDiabetes2` data from the `mlbench` package for predicting the probability of being diabetes positive based on multiple clinical variables. After having loaded the data, edit `??PimaIndiansDiabetes2` for a description of the variables.

1. Randomly split the data into training set (80% for building a predictive model) and test set (20% for evaluating the model). Make sure to set seed for reproducibility.
2. Before fitting a penalized (lasso) logistic regression, find the optimal value of `lambda` that minimizes the cross-validation error.
3. Using `lambda.min` as the best `lambda`, obtain the penalized regression coefficients.
4. Compute the final lasso logistic model, make prediction on test data and calculate the model accuracy.
5. Fit an unpenalized logistic model, make predictions, calculate the model accuracy and compare it with the accuracy of the penalized logistic regression.