

Assignment

For this assignment you will use the dataset “prostate” (prostate2.Rdata). The dataset contains data about prostate cancer patients with information on the size of the prostate, the age of the patient, a blood marker (lpsa) and so on. The response variable is a score (Cscore) on the progression of the cancer after detailed study of the tumor pathology.

1. Study and describe the predictor variables. Do you see any issues that are relevant for making predictions?
2. Make an appropriate LASSO model, with the appropriate link and error function, and evaluate the prediction performance.
3. Look at the coefficient for “lcavol” in your LASSO model. Does this coefficient correspond to how well it can predict Cscore? Explain your observation.
4. Fit a model with *appropriate* non-linear effects. Report a comparison of performance to LASSO and explain what you find.