

Master BeNeFri in Computer Science

Course: Statistical Learning Methods
Spring 2016

Exercise #8. Comparing schemes & Logistic regression

1. Consider dataset `Cars2` dataset (filename: `Cars2Data.txt`, see Exercise #4).

- a) Build three different (generalized) linear regression models (at least one of them must be of the form $\text{mpg} \sim \text{poly}(\text{predictor}, i)$ for some $i \neq 1$).
- b) Perform 10-fold cross validation to estimate the test error of the models you built in a).
- c) Compare the schemes in a) . Can you apply a t-test to see which among the three models performs better on the dataset?

2. Download from folder Exercise#8 on ILIAS website the dataset `Cancer` dataset (filename: `Cancer.txt`) and read the description `CancerDescription.pdf`.

Apply the logistic regression to predict the category `diagnosis` (`malignant` / `benign`) and interpret the most important values of the model you obtain with R.

Can you estimate the error rate of your model?