

Machine Learning 201

Homework 3

Mike Bowles, PhD & Patricia Hoffman, PhD

The South African heart disease data is described on page 122 of the text, "The Elements of Statistical Learning".

This data set can be found on the text book web site: <http://www-stat.stanford.edu/~tibs/ElemStatLearn/>

The sonar data set is on UC Irvine's web site:
<http://archive.ics.uci.edu/ml/machine-learning-databases/undocumented/connectionist-bench/sonar/>

- 1) Use your favorite linear regression technique on the South African data set. What is the least squared error?
- 2) Next use linear discriminant analysis (LDA) on the South African data set. What is the least squared error?
- 3) Now use quadratic discriminant analysis (QDA) on the South African data set. What is the least squared error? Compare the results of the three methods used on the South African data set.
- 4) Use iteratively reweighted least squares (IRLS) unconstrained logistic regression on the sonar data set.
- 5) Finally use regularized logistic regression (glmnet) on the sonar data set. Which of these two methods worked better.
- 6) Look through Joe's data set. Suggest questions about the data that can be answered using techniques that have been covered in class.