Predicting the Age of Abalone Through Regression Analysis

Michael Szczepaniak

October, 2015

Contents

Introduction

A linear regression analysis was performed on a data set of characteristics related to the number of abalone rings using the R analysis package. These characteristics were: Sex (Male, Female, or Infant), Length, Diameter, Height, Whole weight, Shucked weight, Viscera weight, and Shell weight.

The independent variables (characteristics) were first standardized before a matrix of scatter plots were generated to get a high level understanding of the linear relationships between pairs of attributes. From here, we began constructing our linear model by randomly partitioning the data into training and test sets, calculated weights for the linear model for various ridge regression parameters (λ) and training partitions, identified the λ that minimize root mean square error (RMSE) and then used the model to make some predictions.

The parameters for the optimized linear model was as follows: TBD

Background Information

What are Abalone?

TBD

Why Model Rings

TBD

Data Preparation

Regressions Analysis

Experiments

Results

A Personal Curiosity: Bias and Training Partition

Discussion

References