Stats Homework 7

Jon Ashbrock

March 14, 2018

We will use the "NBAStandings1e" and "NBAStandings2016" data sets from Lock5Data in this homework. These data frames contain information about winning percentage, points scored, and points allowed for all NBA teams in two different seasons. No knowledge of basketball is required for this homework other than the team that scores more points is the winner.

You are to build an SVM to classify basketball teams into three categories: "Good", "Average", and "Bad". A team is "Good" if their winning percentage is above 60%, they are average if their winning percentage is between 40% and 60% and bad otherwise. Your homework should:

1. Write code to generate a list (as a factor data-type) which classifies each team as good, average, or bad.

2. Train an SVM on the dataset NBAStandings1e and the list you computed in part 1. Use the points scored and points allowed as your independent variables.

3. Use the SVM to predict whether the teams in 2016 are good, bad, or average.

You should turn in a document summarizing your findings. This document should include at minimum:

1. The confusion tables for the SVM model on each data set with an explanation of what the table tells us

2. A non-mathematical analysis of whether or not you think the model is good based on the confusion table (You can either create a table in the document or include a picture of the table output from **R**)

3. A scatter plot showing the 2016 data. You should color code the points based on their **predicted** category and change the shape of the point based on whether or not it was correctly categorized. What the colors and symbols mean should be explain in the document. Color can be specified in plot by using the "col=" parameter in the plot function. The type of point can be changed using the "pch=" parameter. See: http://www.sthda.com/english/wiki/r-plot-pch-symbols-the-different-point-shapes-available-in-r