Random Forest

Assignment

About the data:

Letâ€™s consider a Company dataset with around 10 variables and 400 records.

The attributes are as follows:

ï‚® Sales -- Unit sales (in thousands) at each location

ï‚® Competitor Price -- Price charged by competitor at each location

ï‚® Income -- Community income level (in thousands of dollars)

ï‚® Advertising -- Local advertising budget for company at each location (in thousands of dollars)

ï‚® Population -- Population size in region (in thousands)

ï‚® Price -- Price company charges for car seats at each site

ï‚® Shelf Location at stores -- A factor with levels Bad, Good and Medium indicating the quality of the shelving location for the car seats at each site

ï‚® Age -- Average age of the local population

ï‚® Education -- Education level at each location

ï‚® Urban -- A factor with levels No and Yes to indicate whether the store is in an urban or rural location

ï‚® US -- A factor with levels No and Yes to indicate whether the store is in the US or not

The company dataset looks like this:

Problem Statement:

A cloth manufacturing company is interested to know about the segment or attributes causes high sale.

Approach - A Random Forest can be built with target variable Sales (we will first convert it in categorical variable) & all other variable will be independent in the analysis.