Lab Exercises

1. Given is the case Bankruptcy given in folder Bankruptcy. The Data contains the following variables:

Name	Description
D	D=0 for bankrupt firms, D=1 for healthy firms.
YR	Year of Bankruptcy for failed firm in matched pair
R1	CASH/CURDEBT
R2	CASH/SALES
R3	CASH/ASSETS
R4	CASH/DEBTS
R5	CFF0/SALES
R6	CFFO/ASSETS
R7	CFFO/DEBTS
R8	COGS/INV
R9	CURASS/CURDEBT
R10	CURASS/SALES
R11	CURRASS/ASSETS
R12	CURDEBT/DEBTS
R13	INC/SALES
R14	INC/ASSETS
R15	INC/DEBTS
R16	UBCDEP/SALES
R17	INCDEP/ASSETS
R18	INCDEP/DEBTS
R19	SALES/REC
R20	SALES/ASSETS
R21	ASSETS/DEBTS
R22	WCFO/SALES
R23	WCFO/ASSETS
R24	WCFO/DEBTS

The variables R1,R2,…R24 are all financial ratios given. We need to find with what accuracy can we predict the Health of the firm based on this information using tree algorithms.

2. The 2008-09 nine-month academic salary for Assistant Professors, Associate Professors and Professors in a college in the U.S. The data Salaries were collected as part of the on-going effort of the college's administration to monitor salary differences between male and female faculty members. The data is available in package car.

   Its data frame with 397 observations on the following 6 variables.

   `rank`

   a factor with levels `AssocProfAsstProf` `Prof`

   `discipline`

   a factor with levels `A` (“theoretical” departments) or `B` (“applied” departments).

   `yrs.since.phd`

   years since PhD.

   `yrs.service`

   years of service.

   `sex`

   a factor with levels `Female` `Male`

   `salary`

   nine-month salary, in dollars.

   Fit a model for predicting regression tree using packages party and rpart. Also generate the measures of accuracy for this model. Which of the model gives more precise result?

3. The dataset Carseats in package ISLR contains a simulated data of sale of car seats.

   Its data frame with 400 observations on the following 11 variables.

   - `Sales`
     - Unit sales (in thousands) at each location
   - `CompPrice`
     - Price charged by competitor at each location
   - `Income`
     - Community income level (in thousands of dollars)
   - `Advertising`
     - Local advertising budget for company at each location (in thousands of dollars)
   - `Population`

  - Population size in region (in thousands)
- `Price`
  - Price company charges for car seats at each site
- `ShelveLoc`
  - A factor with levels `Bad`, `Good` and `Medium` indicating the quality of the shelving location for the car seats at each site
- `Age`
  - Average age of the local population
- `Education`
  - Education level at each location
- `Urban`
  - A factor with levels `No` and `Yes` to indicate whether the store is in an urban or rural location
- `US`
  - A factor with levels `No` and `Yes` to indicate whether the store is in the US or not

Fit a model for predicting variable *sales* using packages party and rpart. Also generate the measures of accuracy for this model. Which of the model gives more precise result?