HOA 732 WIC3 Decision Tree Charles Book Club and Boston Housing Data

Dr. A.K. Singh

1. Charles Book Club is trying to send mail out to book club members to see if they want to buy a book called Art History of Florence. But they don't want to send mail out to all members because most members will not buy the book and they will waste money on mailing. CBC wants to predict who will actually be buying the book.

TABLE 21.1 LIST OF VARIABLES IN CHARLES BOOK CLUB DATASET

Variable Name	Description
Seq#	Sequence number in the partition
ID#	Identification number in the full (unpartitioned) market test dataset
Gender	0 = Male, 1 = Female
M	Monetary—Total money spent on books
R	Recency—Months since last purchase
F	Frequency—Total number of purchases
FirstPurch	Months since first purchase
ChildBks	Number of purchases from the category child books
YouthBks	Number of purchases from the category youth books
CookBks	Number of purchases from the category cookbooks
DoItYBks	Number of purchases from the category do-it-yourself books
RefBks	Number of purchases from the category reference books (atlases, encyclopedias, dictionaries)
ArtBks	Number of purchases from the category art books
GeoBks	Number of purchases from the category geography books
ItalCook	Number of purchases of book title *Secrets of Italian Cooking*
ItalAtlas	Number of purchases of book title *Historical Atlas of Italy*
ItalArt	Number of purchases of book title *Italian Art*
Florence	= 1 if *The Art History of Florence* was bought; = 0 if not
Related Purchase	Number of related books purchased

Predict Florence as a function of the potential predictors.

HOA 732 WIC3 Decision Tree Charles Book Club and Boston Housing Data

Dr. A.K. Singh

2. This dataset Boston Housing Data contains information collected by the U.S Census Service concerning housing in the area of Boston Mass. The data was originally published by Harrison, D. and Rubinfeld, D.L. `*Hedonic prices and the demand for clean air*', J. Environ. Economics & Management, vol.5, 81-102, 1978.

There are 14 attributes in the dataset:

1. CRIM - per capita crime rate by town
2. ZN - proportion of residential land zoned for lots over 25,000 sq.ft.
3. INDUS - proportion of non-retail business acres per town.
4. CHAS - Charles River dummy variable (1 if tract bounds river; 0 otherwise)
5. NOX - nitric oxides concentration (parts per 10 million)
6. RM - average number of rooms per dwelling
7. AGE - proportion of owner-occupied units built prior to 1940
8. DIS - weighted distances to five Boston employment centres
9. RAD - index of accessibility to radial highways
10. TAX - full-value property-tax rate per $10,000
11. PTRATIO - pupil-teacher ratio by town
12. B - 1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town
13. LSTAT - % lower status of the population
14. MEDV - Median value of owner-occupied homes in $1000's

Fit MLR model and Decision Tree Regression Model to medv, and compare the results in terms of Root Mean Square Error (RMSE).