

Data Mining for Success with Mathematica

Bryan M. Minor, Ph.D.
CEO & President,
ScienceOps
12 Oct 2007



Outline

- ScienceOps
- Data mining examples
- Mathematica and RLink
- Model development approach
- Mathematica and R examples
- Conclusion

ScienceOps

- Develops algorithms for private industry
- Work across many areas
- Specialize in using Mathematica
- Industries of emphasis
 - Internet advertising
 - Biotech
 - DOD
 - Gaming theory
 - Data mining

Data Mining

- Nonlinear mixed effects models
- Multivariate adaptive regression splines (MARS)
- Random forests
- Survival/reliability methods
- Bootstrapping
- Classification and regression trees

Data Mining Examples

- Internet advertising
 - Click through probability prediction
 - Sell through probability prediction
- Inventory prediction
 - Follett HEG
- Gaming Theory

R Project

- R is a GNU licensed statistical package
 - <http://www.r-project.org/>
- Research statisticians contribute new analytics to R
- Extensive data analytic/data mining capabilities
- Leveraging work from across statistics community

RLink for Mathematica

- Four main interface functions
 - putR
 - getR
 - REval
 - RGraph
- Supports scalars, vectors, matrices of integers, reals, strings
- Provides most R functionality

Modeling approach

- Development of Model
 - Training set
 - Test set
 - Validation set
- Two state model implementation
 - Run time engine
 - Asynchronous update engine

Conclusions

- Mathematica and R
 - Advanced data mining
 - Leveraging Mathematica
- Accelerate algorithm development
- Avoid reimplementations in other languages with 2 stage model



ScienceOps

- www.scienceops.com
- 425-482-0300
- minor@scienceops.com