TOOLS FOR DISCOVERING PATTERNS IN DATA

EXTRACTING VALUE FROM TABLES, TEXT, AND LINKS

REGISTRATION

- $1350 per person, $950 for each additional person from the same organization
- Includes a complimentary copy of the *Handbook of Statistical Analysis and Data Mining Applications*
- Register at elderresearch.com/training

TESTIMONIALS

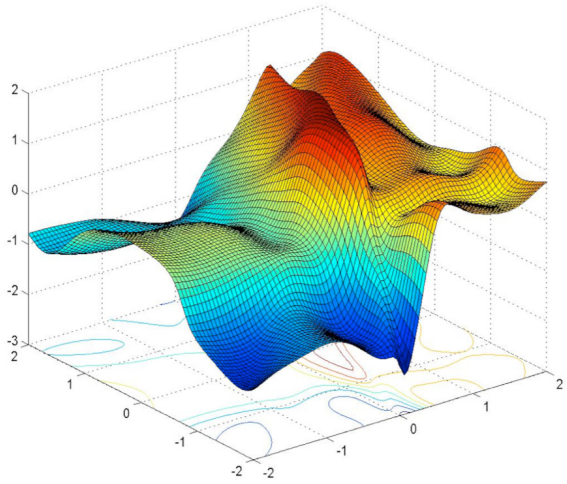
- "Dr. Elder provided examples shedding light on complex concepts. He gave the big picture all along the way."
- "Gave real practical insights from a practitioner's point of view."
- "Finally someone told me how things are done, not just how great Data Mining is."
- "Most valuable, were the insights into the essence of various methods, their relative strengths and weaknesses, and the important open research areas."
- "Very interesting, knowledgeable, and entertaining approach."

Course Description

Find the useful information hidden in your data! This course surveys computer-intensive methods for inductive classification and estimation, drawn from Statistics, Machine Learning, and Data Mining. Dr. Elder will describe the key inner workings of leading algorithms, compare their merits, and (briefly) demonstrate their relative effectiveness on practical applications.

Classical statistical techniques, both linear and nonparametric, will be reviewed and then the ways in which these basic tools are modified and combined into powerful modern methods will be outlined. The course emphasizes practical advice and focuses on the essential techniques of Resampling, Visualization, and Ensembles. Actual scientific and business examples will illustrate proven techniques employed by expert analysts. Along the way, relative strengths and distinctive properties of the leading commercial software products for Data Mining will be discussed.

This intensive short course provides a broad overview, drawing connections between major developments in the diverse fields that contribute to Predictive Analytics, including cutting-edge ways to mine text and graphical networks. Previous participants have found this "big picture" to be very useful for identifying techniques to use immediately, as well as approaches worthy of further exploration, for research or practical problem-solving.



Intended Audience

Those who work with data and wish to understand and use recent developments in predictive analytics. At the conclusion of this course, you should be able to discern the basic strengths of competing methods and select the appropriate tools for your applications.

Contact Us
www.elderresearch.com
contact@elderresearch.com
(434) 973-7673

ELDER RESEARCH
DATA SCIENCE · AI · MACHINE LEARNING

Office Locations
Charlottesville, VA
Washington, DC
Baltimore, MD
Raleigh, NC
London, UK

Course Outline

I. Pattern Discovery: An Overview
- Inducing Models from Data: Benefits and Dangers
- Example Projects from Science and Business
- Characteristics of Successful Projects
- Leading Software Tools and Vendors

II. Classical Statistical Techniques (brief review)
- Regression
- Principle Components
- Nearest Neighbors

III. Modern Methods
- Neural Networks
- Decision Trees

IV. Key General Tools
- Scientific Visualization: Grand Tour, Projection Pursuit, Limitations
- Bootstrapping/Resampling: Essential!
- Optimization: Local and Global
- Target Shuffling: Learning True Significance

V. Data Trouble-Shooting
- Case Diagnostics (Outlying, Influential, Leverage & Missing Points)
- Feature Creation and Selection

VI. Text Mining Overview
- Stemming, Collocation, Feature Engineering
- Statistical vs. Language-dependent Methods
- "Bag of Words" & Vector Space
- Active Learning

VII. Social Network Analysis Overview
- The Power of the "Network Effect"
- Visualization, Modeling Tools, and Examples

VIII. Comparing and Combining Algorithms
- Adaptive Model Structure
- Matching an Algorithm to Your Application
- Experimental Test Results
- Combining Models to Improve Accuracy
- Bagging & Boosting
- Why Ensembles Work

IX. Top 10 Data Mining Mistakes
- Lack Data
- Focus on Training
- Rely on One Technique
- Ask the Wrong Question
- Listen (Only) to the Data
- Leaks from the Future
- Discount Pesky Cases
- Extrapolate
- Answer Every Inquiry
- Sample without Care
- Believe the Best Model

About the Instructor

John F. Elder, IV, Ph.D. heads the US's top data mining consulting team, based in Charlottesville, Virginia, with offices in Washington, DC, Baltimore, MD, and Raleigh, NC. Founded in 1995, Elder Research focuses on commercial, investment, and security applications of advanced analytics, including stock selection, text mining, social networks, image recognition, biometrics, process optimization, drug efficacy, credit scoring, and fraud detection.

John holds Engineering degrees from Rice University, and the University of Virginia, where he's an Adjunct Professor teaching Optimization or Data Mining. Prior to founding Elder Research, he spent a decade in aerospace consulting, investment management, and academia. Dr. Elder has authored innovative data mining tools, is a frequent keynote speaker, and chairs international analytics conferences. He was honored to serve five years on a panel appointed by President Bush to guide technology for national security. He has co-authored award-winning books on practical data mining, ensembles, and text mining.