

TBR
ISM 250: Data Mining and Business Analytics
(Stochastic Optimization in ISTM-I)

Professor Ram Akella

UCSC Silicon Valley Center (Bldg 19-2089 NASA Research Park) & E2-563
akella@soe.ucsc.edu, Tel: 650-279-3078

When: Winter 2008, Tuesdays, 6-9.30 pm (1st Lecture: Tue, Jan 8, 2008)

Where: Silicon Valley Center Bldg 19-2073 and live Telecast to UCSC/Baskin 156

Who (should take this course?): All SOE and SSD-Economics students with strong analytic and business interests including those who wish to develop skills in some of the following areas and obtain attendant benefits:

- Learn about data mining, and analytic/mathematical approaches in business and technology management :
- Search engines and marketing, new product development, supply chain management, financial engineering, and technology and IT management
- explore doing startups in these areas
- start training to work in companies such as Cisco, Yahoo and Google in search and data mining, and HP, IBM, NASA and Agilent in management of technology and business
- Work on projects with Silicon valley firms on these topics
- Explore the possibility of research support based on course and project performance

What (background do you need?): Either possess or learn (in the preliminary boot camp for the course) analytics in linear algebra, probability and statistics; a background in machine learning and/or economics is preferred, though not required

What (should you do?): Register and email your CV to me with your background and interests, including project interests.

Participating Firms: IBM, (SAP)

Course Description

ISM250 and ISM251 form a two course sequence, and is part of the following ISTM/TIM streams:

- Information Retrieval and Knowledge Management Sequence, with 240 (Data Mining) and 260(IRKM) and
- Management of Technology and Services sequence (including 206 (Optimization) and 207 (Stochastic models))

You will learn some of the following skills:

analyze and synthesize business intelligence platform needs at the algorithmic decision making level, including functions such as marketing and sales through data and text mining, new product development speedup, supply chain and e-business optimization, (and possibly financial engineering and risk management).

learn basic analytics of data mining, including supervised and unsupervised learning approaches, search engine and modern information system retrieval

analyze intelligent support systems for marketing decisions, including fundamental methods such as conjoint analysis, together with web search, information retrieval, and data mining approaches, for learning about markets and customer preferences, as well as develop mathematical models for optimizing sales, marketing, and pricing decisions in high tech

learn basic of constrained optimization and dynamic programming, including value and policy iteration for finite horizon situations

perform financial and decision analysis to manage risk and to develop technologies and products that are profitable, by learning financial engineering ideas in portfolios and real options

synthesize and manage an integrated e-business/supply chain network

In addition, projects will be used as a vehicle to better understand and apply the methods.

The course format will be 1/2 lectures, and 1/2 external and internal speakers and seminars including industry personnel, faculty, and students. Significant time will be devoted to project modelling and analysis, and a term (project/research) paper.

The course emphasis will be tuned to the class composition and interest.

COURSE OUTLINE, SYLLABUS, & READINGS

Week 1:

- Business and Management Functions (Innovation – R&D/Product Development/ Engineering, Marketing, Finance/Accounting, Operations/Supply Chain Management, Organizational Management) for Technology Development and Commercialization
- Data/text Mining and Search and role/use in business analytics

Week 2:

- Quantitative market assessment of technology: marketing engineering and optimization
- Data Analysis and metrics/goals in data mining

- Technology and Business Examples: High Technology such as semiconductors, telecommunication, information technology, software, search technology, business analytic technology (supply chain management, new product development)

Week 3

- Constrained Optimization 1 with marketing and product portfolio examples, including pricing
- Advanced Decision Trees in Data Mining 1 (applications in marketing, sales, credit rating, text/document classification, anomalies etc.), including classification metrics

Week 4

- Constrained Optimization 2 (Kuhn-Tucker Conditions) with advanced marketing examples, including web page layout to maximize profits
- Decision Trees in Data Mining 2 – Detailed algorithms (and examples including text mining)

Week 5

- Constrained optimization 3: Applied to Product Portfolios and financial portfolios
- Nearest Neighbour Classifiers - Analytics

Week 6

- Stochastic Dynamic Programming concepts
- Bayesian Classifiers 1 : Naïve Bayes Classifierse

Week 7

- Stochastic Dynamic Programming concepts applied to Real Options in Finance
- Bayes Classifiers and bayesian belief networks

Week 8

- DP in Supply Chain Management and E-Business
- Discriminant functions and Support Vector Machines

Week 9

- Stochastic Optimization in Design Capacity and Risk Management (Waterfall to Spiral Models)
- Artificial Neural Nets, and Ensemble Methods: Bagging and Boosting

Week 10 (plus extra class)

- Association Analysis and Links to Supply Chain Management and E-Business
- Integration of data mining and business analytics, and firm strategy
- Clustering
- Course summary

Entrepreneurship Preparation:

Additional coverage can be provided, based on student interest, by covering Chapters 1-20 of Technology Ventures by Dorf and Byers, in parallel, on a weekly basis

Analytic Bootcamp by TA

Weeks 1-5:

Review of Linear Algebra

Review of statistics

Convex sets and functions

Review of stochastic processes and Markov Chains

Basic Net Present Value Concepts

Software

- Excel
- Matlab
- SAS etc. as needed

Course Grading (May alter to weight project/term/research paper more heavil, if of sufficiently high quality)

Weekly Homework on fundamental topics, quizzes, Comprehensive Course Project/term paper (including presentation to class)

Homework: 20%

Quizzes and final: 30%

Project/Term paper: 50%

Presentation: 10%

Textbooks .

References (An extensive reference list is being provided for course projects and to help with follow course tracks)

Linear Algebra

Linear Algebra, 3rd edition: Strang, Wellesley-Cambridge Press, 2003

Matrix Analysis and Applied Linear : Meyer, SIAM, 2000

Statistics

Analyzing Multivariate Data: Lattin, Carroll, and Green, Thompson, 2003 (Strongly recommended)

Statistical Models: Freedman, Cambridge Press, 2005

Introduction to Probability and Statistics: Ross, Wiley, 1987

Mathematical Programming, Stochastic Models/Processes, and StochasticOptimization

Nonlinear Programming, Mokhtar Bazaraa, and CM Shetty, Wiley, 1979

Nonlinear Programming: Avriel, Dover, 1976/2003

Mathematical Optimization and Economic Theory: Intriligator, SIAM, 1971/2002

Stochastic Processes, Sheldon Ross, Academic Press, 1993

Queueing Systems, Gross and Harris, 1993

Dynamic Programming and Optimal Control, Vols 1- 2, 2nd ed, Dimitri Bertsekas, Athena Scientific, 2000 & 2002

Supply Chain Management and E-Business

E-Business and Supply Chain Networks, Simchi-Levi et.al, Kluwer, 2003

Supply Chain Management: Tayur and Magazine, Kluwer, 1998

Supply Chain Management and E-Business: Management Science Special Issue (Interfaces) 2003

E-Business Management: Ed. By Shaw, Kluwer, 2003 (Strongly recommended)

Data and Text Mining

Pattern Classification, 2nd Edition: Duda, Hart, and Stork, Wiley, 2001 (Strongly recommended)

The Elements of Statistical Learning: Hastie, Tibshirani, and Friedman, Springer, 2001

Data Mining: Witten and Frank, Elsevier/Morgan Kaufmann, 2005

The Search: Battelle, Portfolio, 2005

Principles of Data Mining, David Hand, Heikki Mannila, Padhraic Smyth, Prentice Hall, 2001

Data Mining Techniques: Berry and Linoff, Wiley, 1997

Data Mining: Han and Kamber, Morgan kaufmann, 2001

Data Mining: modeling Data for marketing, Risk, and CRM: Rudd, Wiley, 2001/2003

Understanding Search Engines, Michael Berry and Murray Browne, SIAM, 1999

Modern Information Retrieval, Ricardo Baez-Yates, Berthier Ribeiro-Neto, Addison-Wesley, 1999

AI

Artificial Intelligence., A Modern Approach, 2nd edition: Russell and Norvig, Prentice Hall, 2002

Introduction to Knowledge Systems, Mark Stefik

New Product Development

Management Science: Special Issue on New Product development, 2001

Setting the Pace in New Product Development: McGrath, Elsevier, 1996

Product Leadership: Cooper, Basic Books, 2005

Developing Products in Half the Time, 2nd edition: Smith and Reinertsen, 1998

The Balanced Scorecard: Kaplan and Norton, HBS, 1996

Entrepreneurship

Engineering Your Startup: Baird, Professional Publications, Inc., 1999 (Strongly recommended)

Marketing and/or Search

Search Engine Marketing, Inc.: Moran and Hunt, IBM Press, 2006

Marketing Research, 4th edition: Malhotra, Prentice Hall, 2004

Marketing Management, Phillip Kotler, Prentice Hall, 2002.

Marketing Engineering, Gary Lilien and Arvind Rangaswamy, Prentice Hall, 2003

Marketing Models, Gary Lilien, Philip Kotler, Sridhar Moorthy, Prentice Hall, 1992

Intelligent Support Systems for Marketing Decisions, Nikolaos Matsatsinis and Yannis Siskos, Kluwer, 2003

Finance and Financial Engineering

Investment Science: Luenberger, Oxford University Press, 1998

Principles of Corporate Finance, 7th Edition: Brealey and Myers, McGraw-Hill, 2003

Corporate Finance, 7th edition: Ross, Westerfiled, and Jaffe, McGraw-Hill, 2005

Real Options: Trigeorgis, MIT Press, 1996

Real Options: Amram and Kulatilaka, HBS, 1999

Financial Engineering by Stan Pliska, 2001

Theory of Financial Decision Making: Ingersoll, Rowman & Littlefield, 1987

Financial Modeling in Excel: Benninga, MIT Press, 2001

Strategic Management

Strategic Management: Saloner, Shepard, and Podolny, Wiley, 2001

Prerequisites:

While no formal prerequisites are required, to provide for diverse student backgrounds from AM/BME/CS/CE/Econ/EE/ISTM, the recommended courses include ISM 206, ENGR 203, ENGR 205, CMPE 230 to cover some related aspects. However, in this course, these topics are deepened and focused towards the engineering and management context. Instructor approval based on mathematical maturity is a possibility.

Course Context

This course is intended to be the first in a series of courses in the new Technology and Information Management Program. The motivation for these courses is to teach students the theory and practice of the technology development and management, through the use of information system based decision making. To these ends the courses will integrate information systems, technology and product development, with business issues such as operations, e-business/supply chain management, finance, marketing, and strategy.

This sequence of core courses will form the foundation on which other graduate courses in TIM will be built.

Class Schedule: All Readings must be completed for the class, and all assignments handed in at start of the class, DM= Data Mining, BA= Business Analytics, Ai = Assignmnet #i, Q1, Q2=quizzes, R1= Initial project report, R2= Outline Final report, R3= Final Report, F1 = Final

Lec	Date	Topic	Reading	Assign ment	OP TI ON AL (D B)
1a 1b	1/11/05	BA: Intro DM: Intro	TSK, Ch 1 (Also, The Search, by Battelle)		1,2
2a 2b	1/18/05	BA: Conjoint/mkt DM: Metrics and data	HBR, Conjoint Analysis, A Mgr's Guide, Ch 21, Malhotra TSK, Ch 2	A1	3,4
3a 3b	1/25/05	BA: Constrained Opt and KKT DM: Decision Trees & Metrics	Intriligator, Ch4, Bazaraa & Shetty, Ch4 TSK, Ch 4	A2	5,6
4a 4b	2/1/05	BA: KKT and mkt/product portfolios DM: Decision Trees (continued)	Lilien, Kotler, Moorthy+Readings TSK, Ch 4	A3	7,8
5a 5b	2/8/05	BA: KKT to Financial portfolios DM: Nearest Neighbour	Luenberger Ch 6 (esp.6.6) TSK, Ch. 5.2	A4, Q1, R1	9,1 0
6a 6b	2/15/05	BA: Stochastic Dynamic Programming (SDP) DM: Naïve Bayes Classifiers	Bertsekas Ch 1 TSK Ch. 5.3	A5	11, 12
7a 7b	2/22/05	BA: SDP applied to Real Options DM: Bayes & Belief Networks	Luenberger Ch 12, Trigeorgis, Amram & Kulatilaka, HBR – Luehrman article TSK, Ch 5.3	A6	13, 14
8a 8b	3/1/05	BA: SDP in Supply Chain Management DM: Discriminant functions & Support Vector Machines	Bertsekas Ch 1 TSK, Ch 5.5	A7	15, 16
9a 9b	3/8/05	BA: Opt & SDP in Product Develop & Risk Mgmt DM: Neural nets and Bagging & Boosting	Stanford Note on Variability and Queueing, Management Science, Adler et. al paper TSK, Ch 5.4, 5.6	A8, R2	17, 18
10a 10b (plus 11)	3/15/05	BA: Integration DM: Market Basket Clustering	TSK Ch 6 Ch. 8	A9, Q2	19, 20
	3/18/05			R3, Final F1	

