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 Classifier

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**


**Statistics**

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


**Mathematical Programming, Stochastic Models/Processes, and Stochastic Optimization**

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


Queueing Systems, Gross and Harris, 1993

**Supply Chain Management and E-Business**
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
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**
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 Models, Gary Lilien, Philip Kotler, Sridhar Moorthy, Prentice Hall, 1992

**Finance and Financial Engineering**

Investment Science: Luenberger, Oxford University Press, 1998


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


**Strategic Management**

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

**Prerequisites:**

Wile 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-bsiness/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 = Assignment #i, Q1, Q2= quizzes, R1= Initial project report, R2= Outline Final report, R3= Final Report, F1 = Final

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