FE 582   Page 1

STEVENS INSTITUTE OF TECHNOLOGY

BIA-656: Statistical Learning and Analytics

Syllabus (Spring 2018)

FE582 Instructor: Dragos Bozdog
Office: Babbio 429A
Email: dbozdog@stevens.edu
Phone: (201) 216-3527

Time: Monday (12:00pm-2:30pm)

Room: Hanlon Financial Systems Lab (Babbio 4th floor)

Office Hours: By appointment

Description: The significant amount of corporate information available requires a systematic and analytical approach to select the most important information and anticipate major events. Machine learning algorithms facilitate this process understanding, modeling and forecasting the behavior of major corporate variables.

This course introduces statistical and graphical (machine learning) models used for inference and prediction. The emphasis of the course is in the learning capability of the algorithms and their application to several business areas. The course will combine class presentations, discussions, exercises and case analysis to motivate students and train them in the appropriate use of statistical and econometric techniques.

Objective:
- Learn the fundamental concepts of statistical learning algorithms.
- Explore existent and new applications of statistical learning methods to business problems, and to generic classification problems.
- Learn to solve analytical problems in groups and effectively communicate its results.

Prerequisite BIA-652 Multivariate Data Analysis
or
MGT-620 Statistical Models

Textbooks: No single textbook covers all the topics. Several references will be used and supplementary notes will be provided whenever appropriate.

Main References:
3. CML: Hal Daumé III, A Course in Machine Learning. (link)
Other References:
1. Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, *An Introduction to Statistical Learning with Applications in R*, Springer, 2013 (link)

Papers:
2. Yoav Freund and Llew Mason, The alternating decision tree learning algorithm (link)

Outcomes: By the end of this course, the students will be able to:
1. Understand the foundations of statistical learning algorithms
2. Apply statistical models and analytical methods to several business domains using a statistical language.
3. Recognize the value and also the limits of statistical learning algorithms to solve business problems.
4. Solve a major analytical problem using large and heterogeneous datasets in a group project and communicate its results in a professional way.

Assignments: The assignments must be submitted on Canvas by the deadlines posted on the course website. Each student must submit his/her own report. You should also include the code files if you used a script or wrote a program.

Project: The project requires that participants build a decision support system (DSS) based on one or more methods explored in this course. Each project must be developed by groups of students and they should present a project proposal at the middle of the semester and the final project report at the end of the semester.

Software: R & Python

Grading:
Assignments 60%
Project 40%

Graduate Student Code of Academic Integrity: All Stevens, graduate students promise to be fully truthful and avoid dishonesty, fraud, misrepresentation, and deceit of any type in relation to their academic work. A student's submission of work for academic credit indicates that the work is the student's own. All outside assistance must be acknowledged. Any student who violates this code or who knowingly assists another student in violating this code shall be subject to discipline.

All graduate students are bound to the Graduate Student Code of Academic Integrity by enrollment in graduate coursework at Stevens. It is the responsibility of each graduate student to understand and adhere to the Graduate Student Code of
Academic Integrity. More information including types of violations, the process for handling perceived violations, and types of sanctions can be found at [www.stevens.edu/provost/graduate-academics](http://www.stevens.edu/provost/graduate-academics).

## BIA 656 - Course Schedule (Tentative)

<table>
<thead>
<tr>
<th>Week</th>
<th>Topic</th>
<th>Readings</th>
</tr>
</thead>
</table>
| Week 1| Introduction. MLE, Classifiers via generative models                  | PF {1, 2}
|       |                                                                       | ESL {1, 4.3}
|       |                                                                       | CML {7.5}                 |
| Week 2| Nearest neighbor classifiers, decision trees                          | PF {3}
|       |                                                                       | CML {1, 2.1-2.3, 3}
|       |                                                                       | ESL {2.3, 9.2}            |
| Week 3| Linear classifiers, perceptron, online-to-batch, neural network       | PF {4}
|       |                                                                       | CML {4.7.1,10}
|       |                                                                       | ESL {4.5.1, 11}
|       |                                                                       | Paper1 {1, 2, 3.1, 5}     |
| Week 4| Feature expansions, kernels, SVM                                      | CML {7.7,11}
|       |                                                                       | ESL {4.5.2, 12.2-12.3}    |
| Week 5| Convex losses and ERM, convex optimization, learning theory          | CML {7.2-7.6, 12}
|       |                                                                       | ESL{7.9}                  |
| Week 6| Tail bounds, generalization, cross validation, model performance      | CML {5, 6}, ESL{7.10}     |
|       |                                                                       | PF {5, 7, 8}              |
| Week 7| Reductions, mean variance decomposition, boosting, Bagging, random forests | PF (2)                   |
|       |                                                                       | CML (5.9, 13)             |
|       |                                                                       | ESL (7.1-7.3, 10.1-10.9, 15, 16) |
|       |                                                                       | ADTrees                   |
|       |                                                                       | Bagging                   |
|       |                                                                       | Random Forests            |
| Week 8| Application to finance: algorithmic trading                            | Notes                     |
| Week 9| Linear regression, regularized regression                             | ESL [3.2.0, 3.2.1, 3.2.2, 3.4.1, 3.4.2, 3.4.3] |
| Week 10| K-means clustering, principal component analysis                       | ESL [3.5.1, 13.2.1., 14.3.6, 14.5.1] |
|        |                                                                       | CML {3.4, 15}             |
| Week 11| Relational learning: Bayesian models                                   | PF {9, 11}                |
|        |                                                                       | CML 9                     |
| Week 12| Markov models, hidden Markov models                                   | Rabiner's HMM tutorial {I,II,III} |
| Week 13| Application to marketing: Targeting consumers                         | Case Pilgrim Bank         |
| Week 14| Final Project Presentations                                           |                          |