CSCE 623 Statistical Machine Learning

Spring 2019 Syllabus (page 1) & Schedule (page 2)

# Course Description

# This course introduces statistical machine learning and pattern recognition. Topics include supervised and unsupervised learning for classification, regression and clustering. Techniques include feature selection, feature transformation, linear and nonlinear modeling, as well as model performance assessment and interpretation. The course emphasizes making decisions in the trade space of learning theory, machine learning algorithm design and application.

# Instructor

Dr. Brett Borghetti

Office: Bldg 640, rm 325

Phone: 937-255-3636 ext 4612

[brett.borghetti@afit.edu](mailto:brett.borghetti@afit.edu) — Please start any email subject lines with [CSCE623]

# Meeting Schedule

Lecture: T/R, 0800-1000 Loc: Bldg 646, rm 220

Office Hours: By appointment (contact via email to schedule an appointment).

# Textbook & Resources (R=Required, O=Optional)

(R) Introduction to Statistical Learning (1st ed, corrected 4th printing), by James, Witten, Hastie, and Tibshirani (downloadable book + content: <http://www-bcf.usc.edu/~gareth/ISL/> ) [Our course textbook: “ISLR”]

(O) Python content for ISLR: <https://github.com/JWarmenhoven/ISLR-python> ; <http://sujitpal.blogspot.com/2014/05/ipython-notebooks-for-statlearning.html>

(O) Hands-On Machine Learning with Scikit-Learn & Tensorflow, by Aurelien Geron, Published by ORiley (helpful in CSCE 623 & CSCE 823) [a highly-helpful book on applying machine learning in python: “HOML”]

(O) The Elements of Statistical Learning, by Trevor Hastie, Robert Tibshirani, and Jerome Friedman

(downloadable book/content: <http://statweb.stanford.edu/~tibs/ElemStatLearn/> ) (the math-oriented big brother of ISLR: “ESL”)

# Programming Platform & Packages

Python via Anaconda Installation with Packages: numpy; pandas; scikit-learn; Matplotlib, (Recommended python v3.6 environment or later for compatibility with CSCE 823 – see additional instructions)

Recommended editors: pycharm (for debugging/watching variables) and/or Jupyter (iPython notebook) for interactive programming & reproducible research data/code/documentation products. Homework/Projects will be delivered in a Jupyter notebook.

# Grading

|  |  |
| --- | --- |
| Concept Checks | 8 pts |
| Assignments | 25 pts |
| Final | 30 pts |
| Project & Video | 40 pts |
| **TOTAL** | **103pts** |

Conduct and Ethics: EN OI 36-107  
If you misrepresent other’s work as your own you will receive an ‘F’ for this course, and your case will be reviewed to determine additional academic/UCMJ actions to be taken.

|  |  |
| --- | --- |
| ≥ 97.0 | A |
| [90.0 97.0) | A- |
| [86.0 90,0) | B+ |
| [83.0 86.0) | B |
| [80.0 83.0) | B- |
| [77.0 80.0) | C+ |
| [73.0 77.0) | C |
| [70.0 73.0) | C- |
| [65.0 70.0) | D |
| <65.0 | F |

Learning Outcomes: At the end of the course a student will be able to:

1. Construct a research question which can be answered using the Machine Learning (ML) pipeline.

2. Obtain, clean, and wrangle data to prepare it for use with ML algorithms.

3. Select and apply ML algorithms and use them to fit and tune ML models from data.

4. Assess results and report findings of ML efforts in a reproducible way

Course Schedule on next page…

**Course Schedule**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Week | Date | Class Day | Topic | PRE-Class Activity Due  (concept check complete by 1200 Monday before class) | Assessment (Due by 2359 on this Day) |
| 1 | 26-Mar | 1 | Pretest (ungraded), Rules of Engagement |  |  |
|  | 28-Mar | 2 | Course Overview, Intro to CRISP-DM and KDD | CRISP-DM & KDD reading |  |
| 2 | 2-Apr | 3 | Intro to ML + Statistical Learning part 1 | Ch 1/ Ch 2 &  Linear Algebra Review + Concept Check (Canvas) |  |
|  | 4-Apr | 4 | Statistical Learning part 2 |  |  |
| 3 | 9-Apr | 5 | Linear Regression part 1 | Ch3 + Concept Check (Canvas) |  |
|  | 11-Apr | 6 | Linear Regression part 2 |  |  |
| 4 | 16-Apr | 7 | Classification part 1 | Ch 4 + Concept Check (Canvas) | Asgn 1: Linear Regression |
|  | 18-Apr | 8 | Classification part 2 |  |  |
| 5 | 23-Apr | 9 | Resampling part 1 | Ch 5+ Concept Check (Canvas) | Asgn 2: Classification |
|  | 25-Apr | 10 | Resampling part 2 |  |  |
| 6 | 30-Apr | 11 | TBD - Possible Instructor TDY |  | Asgn 3: Resampling |
|  | 2-May | 12 | TBD - Possible Instructor TDY |  | Project Proposal |
| 7 | 7-May | 13 | Linear Model Selection and Regularization part 1 | Ch 6+ Concept Check (Canvas) |  |
|  | 9-May | 14 | Linear Model Selection and Regularization part 2 |  |  |
| 8 | 14-May | 15 | Tree-Based Methods part 1 | Ch 8 + Concept Check (Canvas) | Asgn 4: Regularization |
|  | 16-May | 16 | Tree-Based Methods part 2 |  | Project First Draft |
| 9 | 21-May | 17 | Support Vector Machines | Ch 9+ Concept Check (Canvas) | Asgn 5: Tree Based Models |
|  | 23-May | 18 | Unsupervised Methods | Ch 10 + Concept Check (Canvas) |  |
| 10 | 28-May | 19 | TBD / wrap up / final exam review |  | Project Video |
|  | 30-May | 20 | FINAL EXAM |  |  |
| 11 – Finals week |  |  | Academy Awards of Machine Learning - Video Viewing Day (date TBD) |  | Project Final Draft  (due 3 June @ 2359) |