# MSA 8150: Machine Learning for Analytics
# Spring 2019

## Course Summary

This course covers a set of topics in machine learning and artificial intelligence, most suitable for the master of science in analytics students. The course covers a variety of topics including the supervised and unsupervised learning, linear and nonlinear regression/classification and boosting techniques. After taking this course, the students are expected to have a good understanding of machine learning basics and able to use the learning tools for real-world and research problems.

## Instructor

- **Instructor:** Alireza Aghasi
- **Office:** Room 542, Buckhead Center
  Room 405B, Downtown Campus
- **Office Hours:** Wednesdays 6:05-7:15p (@ Buckhead Center)
  Also available via a teleconference (see slide notes)
- **Phone:** (404) 413-7718
- **Email:** aaghasi@gsu.edu

## Course Information

- **Location:** GSU Buckhead Center
- **Class Day:** Wednesdays (2:00-4:30p for CRN 19396) and (7:15-9:45p for CRN 15047)
- **Website:** iCollege http://icollege.gsu.edu
- **Teaching Assistant:** TBD (@student.gsu.edu)

- **Prerequisite:** Students should be familiar with the basics of calculus and linear algebra and have had some level of exposure to basic probability. Students should also have some programming experience.
- **Software:** R, Python, MATLAB

# Lecture Notes and Text

Lecture slides/notes along with supplemental materials will be posted on iCollege. The slides do not cover the entire material and the students are advised to take notes during the class, as many topics are taught extensively and interactively via examples.

## Main Text

- James, G., Witten, D., Hastie, T., Tibshirani, R., "An Introduction to Statistical Learning: with Applications in R", Springer, 2013 **(main text)**.
  URL (E-book): https://goo.gl/8NYEo4
- Hastie, T., Tibshirani, R., Friedman, J., "The Elements of Statistical Learning: Data Mining, Inference, and Prediction", Second Edition, Springer, 2009 **(supplemental text)**
  URL (E-book): https://goo.gl/xgr63x

## Additional (Optional) Readings

- Bishop, C., "Pattern Recognition and Machine Learning", Springer, 2006.
  URL: https://goo.gl/56GFVv
- Mitchell, T.M., "Machine Learning", McGraw-Hill, 1997.
  URL: https://goo.gl/HrBDtK

# Grading

There are multiple items that contribute to the final course grade as listed below. There would be **no final exam** and instead a final project is evaluated.

| **Item** | **Percentages** |
|---|---|
| Homework (4-5 Assignments) | 20% |
| In-Class Quizzes (3 quizzes, top 2 counted, each 5%) | 10% |
| Midterm Exam | 25% |
| Knowledge Assessment Test | 5% |
| Final Project | 35% |
| Attendance (see below) | 5% |

The final grade conversion is based on the following table:

| A+ | A | A- | B+ | B | B- | C+ | C | C- | D | F |
|---|---|---|---|---|---|---|---|---|---|---|
| $\geq 97$ | $\geq 90$ | $\geq 87$ | $\geq 83$ | $\geq 80$ | $\geq 77$ | $\geq 73$ | $\geq 70$ | $\geq 67$ | $\geq 60$ | <60 |

## Homework

Homework will be assigned on a biweekly basis (approximately). Homework will be turned in at the beginning of the lecture.

**Late homework is not accepted** and will receive zero credit. You should start working on each homework early, that way you will have enough time to ask questions in the class before the due date. If you cannot make it to the class for any reason, you would need to either submit it electronically before the beginning of the lecture, or ask a classmate to hand it in on your behalf.

Each student must write up and turn in their own solutions. While discussing the homework problems among the students is encouraged, students copying from their classmates or from any other resources will receive a zero score.

Effectively, homework is worth much more than 20% of your grade. The in-class quizzes are often taken immediately after turning in the homework and heavily correlated with the material turned in. It is extremely unlikely that ones does well on the exam/quizzes without putting enough effort into the homework.

## In-Class Quizzes

There will be 3 in-class quizzes. The date of the quiz will be announced ahead of the time. Most likely, they will be immediately after turning in a homework assignment and heavily correlated with the content of the homework. Among the 3 quizzes the top 2 will counted towards your grade. Note that **there will be no make-up for the quizzes**.

## Exams

There will be a **midterm exam** as listed in the class schedule below. The midterm exam covers 25% of the course grade. Also close to the end of the semester, there will be a **knowledge assessment test** (covering 5% of the course grade), which evaluates student's general knowledge of machine learning. This is a short test compared to the midterm, but covers the majority of the material presented in the course.

The exams and all quizzes are closed-book and closed-notes. There would be no actual final exam for this course and a final project replaces it.

Unless you are explicitly told that a calculator is allowed on a quiz or exam, there should not be any calculator within your reach during a quiz or an exam. No communication device, such as a mobile phone or a device that facilitates access to the internet, may be within your reach during a quiz or an exam.

Make-up exams only apply to the midterm and will be scheduled **only** in case of unavoidable emergencies and after the department's approval. Personal business, such as interviews and travel arrangements do not warrant a make-up exam or an incomplete grade. In case of an unavoidable emergency affecting your exam, you should contact the instructor, immediately.

### Final Project

Few weeks after the midterm, students will receive a document presenting the final project guidelines. The final project is normally held as a competition among student teams, and involves addressing a comprehensive set of problems, written reports and in-class presentation.

## Attendance Policy

**Lecture attendance is mandatory and will count towards your grade**. A sign in sheet (starting in week 2) will be passed around at every lecture; please sign next to your name (only).

## Academic Honor Code

All course participants are expected and required to abide by the Georgia State University Honor Code. For details, see the University code of conduct:
https://codeofconduct.gsu.edu/.
Please familiarize yourself with the code, and use it to guide your conduct. Specifically, you must do your own work in all the homework assignments, quizzes and exams. Any form of academic dishonesty, such as plagiarism, can result in serious consequences and will be reported to the department administration.

## Lectures Outline

The outline below should be treated as an approximation; it is subject to some small changes and there would be some overlap among the sessions.

| **Date** | **Topic** | **Reading** |
|---|---|---|
| Jan. 16 | Introduction to learning theory and fundamental notions | ISL.1-2 |
| Jan. 23 | Linear regression and extensions | ISL.3, ESL.3 |
| Jan. 30 | Classification (logistic regression, Naive Bayes, LDA, QDA) | ISL.4, ESL.4 |
| Feb. 6 | Boosting and Cross Validation | ISL.5, ESL.7,8 |
| Feb. 13 | Model selection and regularization (subsets, PCR. PLS, LASSO, Ridge) | ISL.6, ESL.7 |
| Feb. 20 | More on model selection | ISL.6, ESL.7,10 |
| Feb. 27 | Beyond linearity (local smoothers: polynomials, kernel, Splines, GAMs) | ISL.7, ESL.6 |
| Mar. 6 | Overview of linear regression and classification tools | ISL1-7 |
| Mar. 13 | Midterm exam | |
| Mar. 20 | Spring break (no class) | |
| Mar. 27 | Tree-based methods (random forests, decision trees, ...) | ISL.8, ESL.9 |
| Apr. 3 | Nonlinear learning modules, neural networks, deep learning | ESL.11 |
| Apr. 10 | Unsupervised learning (PCA, K-means, ...) | ISL.10 |
| Apr. 17 | Support vector machines | ISL.9, ESL.12 |
| Apr. 24 | Class summary + final presentations (partial) | ESL.11+ML Lit |
| May. 1 | Final presentations (remaining groups) | |