CS 6980: Introduction to Data Science Spring 2018

Syllabus

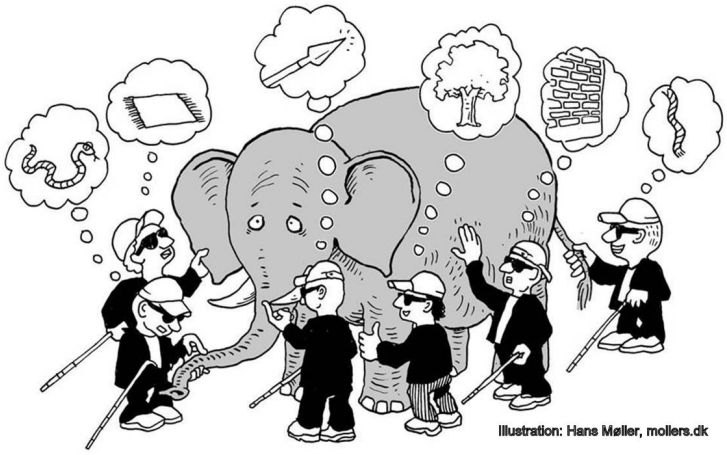
Instructor: Daniel L. Pimentel-Alarcón

Course Description

This course provides a high-level, hands-on introduction to several popular but basic machine learning algorithms, exposing students to modern real-life problems of data science. This should provide students with the necessary background for more advanced data science courses offered in our department, such as machine learning, data mining, big data programming, etc.

Instructor

Daniel L. Pimentel-Alarcón
email: pimentel@gsu.edu
office: 25 Park Place, Suite 742
website: `http://danielpimentel.github.io`



As we move along the course, we will dive into some math. We will see some equations, some theorems, some algorithms, etc. Knowing your math is important. It is important to understand the steps of an equation or the proof of a lemma. Each of these is an important **piece to solve a problem**. However, be careful not to loose track of the problem! It is equally important (or perhaps even more so) to **understand the big picture** and how the pieces come together.

Lectures

Monday, Wednesday
10:00am-11:45am
Classroom South 306

Office Hours

Monday, 8:45am-9:45am,
Wednesday 11:45am-12:45pm.

Prerequisites

Basic background in:

- Linear Algebra
- Probability
- Statistics
- Programming

Grading

- 40% Mini-Projects
- 30% Scribing
- 20% Midterm exam
- 10% Final exam*

* Your final grade will be the **minimum** between your final exam grade and your weighted grade.

Topics

- Simulations & Data Visualization
- Linear Regression
- Logistic Regression
- Gradient Descent
- Nearest Neighbors

- K-Means Clustering
- Information Theory
- Decision Trees & Random Forests
- Expectation & Estimation
- Covariance
- Vector Spaces & Principal Component Analysis (PCA) (if time allows)
- Neural Networks (if time allows)

Lecture notes and Scribing

The content of this course, as well as the latex template for scribing can be found at:

`https://danielpimentel.github.io/teaching.html`

Mini-Projects

Mini-projects related to:

- Diabetes prediction (Visualization and Simulation)
- Disaster survival (Regression)
- Breast cancer (K-means clustering)
- Entropy (Information theory)
- Brain imaging (Statistical learning)
- Face clustering (Principal component analysis)
- Digits classification (Neural networks)

Details will be posted at:

`https://danielpimentel.github.io/teaching.html`

Additional resources

- An Introduction to Statistical Learning with Applications in R (James, Witten, Hastie, and Tibshirani), available at: `http://www-bcf.usc.edu/~gareth/ISL/`.
- Neural Networks and Deep Learning (M. Nielsen), available at: `http://neuralnetworksanddeeplearning.com/index.html`
- Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies (Kelleher, Mac Namee, and D'Arcy).
- Data Mining: Concepts and Techniques (Han, Kamber, and Pei).

Remarks

- The course syllabus provides a general plan for the course; deviations may be necessary.
- Your constructive assessment of this course plays an indispensable role in shaping education at Georgia State. Upon completing the course, please take the time to fill out the online course evaluation.
- Students who wish to request accommodation for a disability may do so by registering with the Office of Disability Services. Students may only be accommodated upon issuance by the Office of Disability Services of a signed Accommodation Plan and are responsible for providing a copy of that plan to instructors of all classes in which accommodations are sought.
- Students are strongly encouraged to work together on homework assignments, but each student must submit his or her own writeup. Plagiarism of material written by classmates, book or article authors, or web posters is prohibited. Students must work independently on exams. Academic integrity will be strictly enforced.