ECE 398BD: Making Sense of Big Data

Fall 2018

http://courses.engr.illinois.edu/ece398BD

Instructors: Lav R. Varshney, Minh N. Do, and Subhonmesh Bose

Teaching Assistants: Yuheng Bu, Vaishnavi Subramanian, and Khaled Alshehri

Course Coordinator: Minh Do

Prerequisites: ECE 313 (or campus equivalent on basic undergrad probability) and some basic linear algebra. General mathematical maturity expected of engineering undergraduates.

Target Audience: Juniors or Seniors

Outline:
Data science has emerged as a key discipline within the mathematical side of engineering, and is applicable in nearly all topical domains. Moreover, big data is all around us: Petabytes of data are collected by Google and Facebook, 24 hours of video are uploaded on Youtube every minute. Making sense of all this data in the relevant context is a critical question. This course takes a holistic view towards understanding how this data is collected, represented and stored, retrieved and computed/analyzed upon to finally arrive at appropriate outcomes for the underlying context. The course is based around three topics, drawn from diverse areas of engineering. Topics for the Fall 2018 offering are as follows.

- Machine learning forms the basis of many predictive big data algorithms. In this topic, we study the basic tools of machine learning, including classification, regression, clustering and dimensionality reduction. Each topic will be illustrated with simple real data sets, and should help students apply the correct statistical tools to other data sets such as those encountered in the remainder of the course.
- Audio and video data are widely available online. For example, camera phones that generate millions of pixels in milliseconds are carried around all the time by billions of people worldwide. Surveillance cameras in a typical company site generate about terabytes of video every day. These ubiquitous visual recording devices generate big unstructured data that provide gold mine for analytics. In this part of the course, students will learn how these data types are acquired, sampled, and stored. Concrete analytics problem involving audio recognition (similar to the commercially available Shazam software) and object detection and monitoring system will be studied.
- The vision to create a smart grid relies on utilizing available data to make smarter decisions in power system operation. In this part of the course, we will explore where machine learning algorithms can help in making such decisions. Using the popular toolkit TensorFlow, we will study prediction, classification, and controller design tasks that arise in grid operation, and understand the practical implications of the performance of such algorithms.

Each of these topics is covered at four discrete levels.

- We start with the context of where the data comes from, how it is acquired, what are the biases and noise levels in the data leading to statistical and physical models of the data acquired.

- Appropriate data representation mechanisms and distributed storage and computing architectures are discussed next. Based on the type of the data, different compression/coding methods are appropriate. Images, videos, social network data, medical imaging data, and smart grid data, each bring their own unique characteristics which can be harnessed towards efficient representation.
- Once data is stored and represented efficiently, we look for the right statistical and algorithmic tools to analyze the data.
- Finally, the analyzed data leads to appropriate inferences or visualizations as appropriate to the engineering problem we started out with. This closes the loop bringing utility to the original setting and context in which the data was acquired.

The idea is that each of the three topics/stories will be taken through the four cycles highlighted above. They are distinct enough to bring about different aspects of the four issues but similar in the sense that the overall plot of what is being covered is similar. An important component of this course is that it is project based. The four topics will be covered sequentially, each ending with a project and a concrete (software) laboratory outcome. The technical material needed is developed on an as-needed-basis, drawing from previous courses (prerequisites) and online material (corequisites). The four topics are synchronized in the sense that the technical material is streamline and once covered, will be amortized in later parts of the course.

Course Plan

Part 1 (Weeks 1-5): Fundamentals of Machine Learning by Lav Varshney

Lecture 1: Introduction to the course; Review of linear algebra and probability
Lecture 2: Elements of Machine Learning
Lecture 3: k-Nearest Neighbor Classifier and Bayes Classifier
Lecture 4: Preview of Parts 2 and 3 [Do and Bose]
Lecture 5: Linear Classifiers and Linear Discriminant Analysis
Lecture 6: Kernel Tricks and Support Vector Machines
Lecture 7: Linear Regression
Lecture 8: K-means Clustering
Lecture 9: SVD and Eigen-Decomposition
Lecture 10: PCA and Applications

Labs (Weeks 1-5)
Lab 1: Introduction to Python
Lab 2: Linear Classification
Lab 3: Kernel Tricks and SVMs
Lab 4: Linear Regression and Clustering
Lab 5: Principal Components

Grading: 30% quizzes (at end of Thursday lectures), 70% labs and lab reports

Part 2 (Weeks 6-10): Audio and Visual Analytics by Minh N. Do

Lecture 1: Introduction to audio and visual analytics. Example applications.
Lecture 2: Signal acquisition and sampling. Audio and visual sensors.
Lecture 3: Audio spectral analysis: DFT, short-time Fourier transform
Lecture 4: Audio content identification. Example: Shazam system

Lecture 5: Visual global feature extraction: color histograms
Lecture 6: Visual local feature extraction: keypoint detection and description, SIFT
Lecture 7: Image search: query by example using color histogram and feature matching
Lecture 8: Video analytics: background modeling and subtraction
Lecture 9: Video analytics: motion detection and tracking
Lecture 10: Introduction to deep learning for audio and visual recognition

Labs (Weeks 6-10)
Lab 1: Audio and visual acquisition, inspection, and visualization in Python and OpenCV
Lab 2: Spectral analysis; k-means clustering and classification of spectral feature vectors
Lab 3: Audio content identification; implement the Shazam algorithm
Lab 4: Image feature extraction in OpenCV/Python. Feature matching and visual search
Lab 5: Background subtraction; motion detection and tracking in video

Grading: 30% quizzes (in lectures), 70% labs and lab reports

<u>Part 3 (Weeks 11-15): Smart Grid by Subhonmesh Bose</u>

Lecture 1: Introduction to power system and the need for load prediction
Lecture 2: Effect of renewable generation on scheduling decisions
Lecture 3: Power flow equations and role of sensor measurements in system operation
Lecture 4: Adversarial tampering v/s erroneous sensors
Lecture 5: Frequency dynamics and the role of control design
Lecture 6: Maintaining grid frequency using neural networks
Lecture 7: The increasing role of smart meters
Lecture 8: Understanding your energy usage from smart meter data
Lecture 9: TBD
Lecture 10: TBD

Labs (Weeks 11-15)
Lab 1: Predicting California's power demands
Lab 2: Detecting bad sensors in power system monitoring
Lab 3: Controlling grid frequency at 60Hz
Lab 4: End-consumers load disaggregation
Lab 5: TBD

Grading: 30% quizzes (at end of Thursday lectures), 70% labs and lab reports

Run > Hide > Fight

Emergencies can happen anywhere and at any time. It is important that we take a minute to prepare for a situation in which our safety or even our lives could depend on our ability to react quickly. When we're faced with almost any kind of emergency – like severe weather or if someone is trying to hurt you – we have three options: Run, hide or fight.

Run

Leaving the area quickly is the best option if it is safe to do so.

- Take time now to learn the different ways to leave your building.
- Leave personal items behind.
- Assist those who need help, but consider whether doing so puts yourself at risk.
- Alert authorities of the emergency when it is safe to do so.

Hide

When you can't or don't want to run, take shelter indoors.

- Take time now to learn different ways to seek shelter in your building.
- If severe weather is imminent, go to the nearest indoor storm refuge area.
- If someone is trying to hurt you and you can't evacuate, get to a place where you can't be seen, lock or barricade your area if possible, silence your phone, don't make any noise and don't come out until you receive an Illini-Alert indicating it is safe to do so.

Fight

As a last resort, you may need to fight to increase your chances of survival.

- Think about what kind of common items are in your area which you can use to defend yourself.
- Team up with others to fight if the situation allows.
- Mentally prepare yourself – you may be in a fight for your life.

Please be aware of people with disabilities who may need additional assistance in emergency situations.

Other resources

- **police.illinois.edu/safe** for more information on how to prepare for emergencies, including how to run, hide or fight and building floor plans that can show you safe areas.

- **emergency.illinois.edu** to sign up for Illini-Alert text messages.

- **Follow the University of Illinois Police Department** on Twitter and Facebook to get regular updates about campus safety.