Course Syllabus – Winter 2018

Instructor: David Boy
Classroom: State Hall Room 206
Office Hours: Online, by appointment

Web Sites
\[ \text{http://blackboard.wayne.edu} \]

Description: This course will cover some fundamentals of statistical methods and techniques required for data science. We will use various statistical learning techniques as a framework for studying the life-cycle of a statistics-driven data science and analytics project through case studies, hands-on work, and a final project. We will work with a variety of software tools and technologies for handling large datasets.

The course will be loosely broken up into three parts.

Part one will focus on building a foundation for talking about statistics, machine learning, and data science. We will learn about Bayesian and classical statistics. We will review and extend our understanding of hypothesis testing. We’ll discuss distributed computing and how it’s special. We’ll touch on some key topics in statistics, such as the bias-variance tradeoff. Also, we will get up-to-speed with the software tools we’ll be using throughout the course.

The second part will focus on applying statistical tools. We will look at regression, classification, feature extraction, dimensionality reduction, and more. Our emphasis will be on learning
- how to build statistical models,
- how to evaluate their usefulness and shortcomings, and
- how to use them to drive good decisions and make meaningful predictions.

The third part will address special topics. This will be guided largely by student interest. Possible topics include natural language processing, collaborative filtering, and time-series analysis. We will also address advanced topics related to student projects.

The class will meet twice per week. Our plan is to use one day for student-led presentation of topics from the previous week and, and to introduce new topics for the following coming week. The other day will focus on examples, case-studies, and hands-on applications. Find a calendar of topics at the end of this document.

\[ ^1 \text{Blackboard website is protected by individual user login names and passwords. The username is the uniquely assigned WSU AccessID. To activate your WSU AccessID or change the password or set an alternate forwarding e-mail address, visit} \text{https://computing.wayne.edu/accessid.} \text{Call the WSU Computing & Information Technology (C&IT) Help Desk at 313-577-4778 for any difficulties.} \]
Course Learning Outcomes: At the end of the course, a successful student should be able to

- Use statistical concepts to describe and understand data and how it does and does not represent the real world.
- Apply statistical methods to data in order to test hypotheses or make predictions.
- Design and execute analyses that lead to insight, decisions, and actions.
- Communicate the results and conclusions of quantitative analyses.
- Learn from and critique statistical analyses done by others.

Prerequisites: Knowledge of a programming language.

Resources: 


Other resources will be used as needed.

Software: In this course, we use R and IBM cloud computing. Using a different programming language and other computing resources will be allowed, but will not be supported by the instructor.

Grading: The course will consist of

- weekly assignments and in-class presentations,
- a project,
- a final exam.

Each will be worth 100% of the course. That is to say, extraordinary work on any of these aspects of the course is sufficient to receive an “A” in the course. Details and expectations shall be explained on the first day of class.

Assignments: Students are encouraged to bring some datasets/problems of interest. If requested, instructor can provide datasets.

Semester Project: This is a team project with two to three students. Students will analyze a data set of interest, develop models, and produce a report.

Attendance Policy: Students attending any given class are required to join the class within the first five minutes to minimize any class disruptions.

Religious Holidays: Because of the extraordinary variety of religious affiliations of the University student body and staff, the Academic Calendar makes no provisions for religious holidays. However, it is University policy to respect the faith and religious obligations of the individual. Students with classes or examinations that conflict with their religious observances are expected to notify their instructors well in advance so that mutually agreeable alternatives may be worked out.

Student Services: The Academic Success Center (1600 Undergraduate Library) assists students with content in select courses and in strengthening study skills. Visit http://success.wayne.edu for
schedules and information on study skills workshops, tutoring and supplemental instruction (primarily in 1000 and 2000 level courses).

The Writing Center is located on the 2nd floor of the Undergraduate Library and provides individual tutoring consultations free of charge. Visit http://clasweb.clas.wayne.edu/writing to obtain information on tutors, appointments, and the type of help they can provide.

**Class Recordings:**

Students need prior written permission from the instructor before recording any portion of this class. If permission is granted, the audio and/or video recording is to be used only for the student’s personal instructional use. Such recordings are not intended for a wider public audience, such as postings to the internet or sharing with others. Students registered with Student Disabilities Services (SDS) who wish to record class materials must present their specific accommodation to the instructor, who will subsequently comply with the request unless there is some specific reason why s/he cannot, such as discussion of confidential or protected information.

**Academic Dishonesty – Plagiarism and Cheating:**

Academic misbehavior means any activity that tends to compromise the academic integrity of the institution or subvert the education process. All forms of academic misbehavior are prohibited at Wayne State University, as outlined in the Student Code of Conduct (http://www.doso.wayne.edu/student-conduct-services.html). Students who commit or assist in committing dishonest acts are subject to downgrading (to a failing grade for the test, paper, or other course-related activity in question, or for the entire course) and/or additional sanctions as described in the Student Code of Conduct.

- **Cheating:** Intentionally using or attempting to use, or intentionally providing or attempting to provide, unauthorized materials, information or assistance in any academic exercise. Examples include: (a) copying from another student’s test paper; (b) allowing another student to copy from a test paper; (c) using unauthorized material such as a “cheat sheet” during an exam.
- **Fabrication:** Intentional and unauthorized falsification of any information or citation. Examples include: (a) citation of information not taken from the source indicated; (b) listing sources in a bibliography not used in a research paper.
- **Plagiarism:** To take and use another’s words or ideas as one’s own. Examples include: (a) failure to use appropriate referencing when using the words or ideas of other persons; (b) altering the language, paraphrasing, omitting, rearranging, or forming new combinations of words in an attempt to make the thoughts of another appear as your own.

Other forms of academic misbehavior include, but are not limited to: (a) unauthorized use of resources, or any attempt to limit another student’s access to educational resources, or any attempt to alter equipment so as to lead to an incorrect answer for subsequent users; (b) enlisting the assistance of a substitute in the taking of examinations; (c) violating course rules as defined in the course syllabus or other written information provided to the student; (d) selling, buying or stealing all or part of an un-administered test or answers to the test; (e) changing or altering a grade on a test or other academic grade records.

**Student Disability Services:**

If you have a documented disability that requires accommodations, you will need to register with Student Disability Services for coordination of your academic accommodations. The Student Disability Services (SDS) office is located at 1600 David Adamany Undergraduate Library in the Student Academic Success Services department. SDS telephone number is 313-577-1851 or 313-577-3365 (TTD only). Once you have your
accommodations in place, I will be glad to meet with you privately during my office hours or at another agreed upon time to discuss your needs. Student Disability Services’ mission is to assist the university in creating an accessible community where students with disabilities have an equal opportunity to fully participate in their educational experience at Wayne State University. Please refer to the SDS website for further information about students with disabilities and the services we provide for faculty and students: http://studentdisability.wayne.edu/

Students who are registered with Student Disability Services and who are eligible for alternate testing accommodations such as extended test time and/or a distraction-reduced environment should present the required test permit to the professor at least one week in advance of the exam. Federal law requires that a student registered with SDS is entitled to the reasonable accommodations specified in the student’s accommodation letter, which might include allowing the student to take the final exam on a day different than the rest of the class.

**Course Drops and Withdrawals:** In the first two weeks of the (full) term, students can drop this class and receive 100% tuition and course fee cancellation. After the end of the second week there is no tuition or fee cancellation. Students who wish to withdraw from the class can initiate a withdrawal request on Pipeline. You will receive a transcript notation of WP (passing), WF (failing), or WN (no graded work) at the time of withdrawal. No withdrawals can be initiated after the end of the tenth week. Students enrolled in the 10th week and beyond will receive a grade. Because withdrawing from courses may have negative academic and financial consequences, students considering course withdrawal should make sure they fully understand all the consequences before taking this step. More information on this can be found at: http://reg.wayne.edu/pdf-policies/students.pdf

**Deferred Grade:** A grade of 'I' can only be assigned if all of the following criteria are met:
1. the student IS NOT currently failing the class and,
2. there is NOT a substantial quantity of work yet to be completed,
3. there is no extra work required of the instructor beyond the normal duties of grading the paper/exam,
4. there is no need for the student to attend the class in subsequent terms.
The final decision to assign an incomplete grade rests with the instructor. An 'I' grade MUST be made up within one year of assignment of the grade.

**Schedule of topics**

<table>
<thead>
<tr>
<th>Week</th>
<th>Main topic(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Distributed computing, cloud computing, and data science ecosystems</td>
</tr>
<tr>
<td>2</td>
<td>Data representation in computing, in statistics, and in machine learning</td>
</tr>
<tr>
<td>3</td>
<td>Distributions, metrics, and regression</td>
</tr>
<tr>
<td>4</td>
<td>Dimensionality reduction</td>
</tr>
<tr>
<td>5</td>
<td>Hypothesis testing</td>
</tr>
<tr>
<td></td>
<td>Classification</td>
</tr>
<tr>
<td>---</td>
<td>----------------</td>
</tr>
<tr>
<td>7</td>
<td>Bayesian and classical statistics</td>
</tr>
<tr>
<td>8</td>
<td>Naive Bayes (or neural networks)</td>
</tr>
<tr>
<td>9</td>
<td>Decision trees</td>
</tr>
<tr>
<td>10</td>
<td>Clustering</td>
</tr>
<tr>
<td>11</td>
<td>Project definition and special topics (see below)</td>
</tr>
<tr>
<td>12</td>
<td>Special topics (see below)</td>
</tr>
<tr>
<td>13</td>
<td>Special topics (see below)</td>
</tr>
<tr>
<td>14</td>
<td>Projects and review</td>
</tr>
</tbody>
</table>

**Special topics**
Which topics the course addresses will depend on student and instructor interest, to be determined.

- NLP and text-mining
- Bayesian estimation (supersedes the t-test)
- time-series analysis
- gaussian mixture models
- collaborative filtering
- pattern matching
- neural networks
- more?