

WAYNE STATE UNIVERSITY

COLLEGE OF ENGINEERING

DSA 6100: Statistical Methods for Data Science and Analytics Course Syllabus - Winter 2019

- Instructor:** Qingyu Yang, Ph.D.
Associate Professor, Industrial & Systems Engineering Department
qyang@wayne.edu | Tel: (313) 577- 9665
- Teaching Assistant:** Aminisharifabad Mohammad, Ph.D. Student, Office location: ISE Building Room 1062.
mohammad.aminisharifabad@wayne.edu
- Time & Location:** Monday & Wednesday 7:00 – 8:15 pm, Room 1137, Old Main (MAIN), 4841 Cass Ave.
- Office Hours:** Monday 4:20-5:20 pm, Or by Appointment, Room 2167, 4815 Fourth Street (MEB), Wayne State University, Detroit, MI 48202
- Web Sites:** <http://canvas.wayne.edu>
- Description:** A fundamental course covering statistical methods and techniques required for data science and analytics applications through methods, case studies, and a semester project that cuts across all course modules. This course focuses on both statistical methods and the life-cycle of a statistics driven data science and analytics project. Students will be exposed to a variety of tools and technologies. In many cases, a parsimonious and interpretable model is favored over a black-box type learning algorithm, even when it means that one has to sacrifice predictive accuracy to some extent. How to properly formulate a statistical model, how to develop intuitive insights and interpretation, and how to evaluate the uncertainty of the analysis are emphasized in this course with algorithmic details.
- Course Learning Outcomes:** Upon successful completion of the course, learning partners will be able to:
- Apply the concepts and procedures of inferential statistics including probability, statistical models, and hypothesis testing for data science analysis
 - Use statistical software to conduct a variety of statistical analyses, including testing of statistical assumptions, data transformations, and validation of statistical findings
 - Apply the statistical learning and data analysis strategy to solve real-world problems and test hypothesis, with specifications for data elements, requirements of the statistic, and limitations to the interpretation
 - Discuss results of statistical analyses efficiently, and prepare written reports and technical illustrations that summarize background, analysis procedures, and interpretation of technical results
- Students will also be exposed to emerging trends, new software tools and best business practices in the DSA area via homework assignments and case studies
- Prerequisites:** Basic statistical knowledge (e.g., hypothesis tests, confidence intervals, p-values, as well as the use and interpretation of statistical model results). Familiarity with statistical programming language R.
- Textbook and References:**
- James, Witten, Hastie and Tibshirani, *An Introduction to Statistical Learning with Applications in R*. [[Free PDF](#)], .
 - Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*. New York: Springer, [[Free PDF](#)], 2001
 - Y. Abu-Mostafa, M. Magdon-Ismael, and H. Lin., *Learning from Data: A Short Course*, AMLBook, 2012.
- Additional tutorials and journal papers will be distributed during the class as needed to complement the material from the textbooks.
- Software:** RStudio, SParkR, Hadoop

Grading:	Homework 10%
	Final Exam 20%
	Critique of Case studies 15%
	Semester Project 40%
	Class Discussion 5%
Homework Policy:	Homework assignments and solution will be posted on course website and announced in class. Homework will be due at the beginning of the class. Homework submitted late by one class will be evaluated at 50% credit and late by two (or more) classes will not receive any credit. In case the student is unable to attend class for submitting the homework, e-mail the work to the Teaching Assistant.
Exams:	Exams will be open-book open-note, focusing on the material taught in class. Makeup examinations must be requested in writing and must be the direct result of a medical or work-related issue.
Semester Project:	During the semester, a course project is required by applying the statistical methods to solve an interesting real-world big data problem. Detailed guideline of the project will be given during the semester. The project will be graded as follows: <ul style="list-style-type: none"> • Proposal 5% • Data Set Collection/Pre-analysis 10% • Presentation 35% • Final Project Report 50%
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:	<ul style="list-style-type: none"> ○ <i>The Academic Success Center</i> (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). ○ <i>The Writing Center</i> 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. <ul style="list-style-type: none"> ○ <u><i>Cheating</i></u>: 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. ○ <u><i>Fabrication</i></u>: 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. ○ <u><i>Plagiarism</i></u>: 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. The SDS telephone number is 313-577-1851 or 313-202-4216 for videophone use. Once you have met with your disability specialist, I will be glad to meet with you privately during my office hours to discuss your accommodations. 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. You can learn more about the disability office at <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.

Tentative Course Schedule:

Module	Topics	Homework and Assignment
Module 1	Introduction <ul style="list-style-type: none"> - Introduction of class - Introduction of statistical learning - Introduction of R 	
Module 2	Ordinary Least Squares Linear Regression <ul style="list-style-type: none"> - Simple and multiple linear regression - Ordinary least squares estimation for linear regression models. 	
Module 3	Linear model for high dimensional data <ul style="list-style-type: none"> - Model selection - Ridge regression - LASSO 	Modeling assignment 1
Module 4	Dimension reduction <ul style="list-style-type: none"> - Multi-dimensional scaling - Principal components analysis - Supervised dimension reduction 	Project proposal due
Module 5	Resampling* (*depends on time) <ul style="list-style-type: none"> - leave-one-out cross-validation - K-fold cross-validation. - Bootstrap method 	Modeling assignment 1 due Modeling assignment 2
Module 6	Nonlinear models <ul style="list-style-type: none"> - Polynomial regression - Step functions - Regression and smoothing splines - Local regression - Generalized additive models (GAMs) 	
Module 7	Classification <ul style="list-style-type: none"> - Bayes rule - Binary logistic regression algorithm. - Multi-class logistic regression algorithm. - Linear, quadratic, and mixture discriminant analysis - Review of nearest neighbor classifier 	Modeling assignment 2 due Modeling assignment 3
Module 8	Support Vector Machine <ul style="list-style-type: none"> - Support Vector Machine - Kernel Based method 	Critique report due
Module 9	Unsupervised learning <ul style="list-style-type: none"> - k-means - hierarchical clustering - model-based clustering 	Modeling assignment 3 due
Module 10	Artificial neural network* (*depends on time) <ul style="list-style-type: none"> - Perceptron learning algorithm - Artificial neural networks - Back-propagation algorithm 	
Module 11	Statistical methods for big data (depends on time) <ul style="list-style-type: none"> - Hadoop - SparkR 	Final exam Project report dues

**Potential Topics
depending on
Time**

- Time series
- Network and graphical models
- NLP and text-mining
- Spatial statistics

**Grading
Scale:**

The grade of homework assignment, exam, and project might be curved. The final grade will be assigned according to the following numeric scores obtained from homework assignments, final exam, and the course project report/presentation:

Course Points (0-100)	Course Grade
91 – 100	A
86 – 90	A-
81 – 85	B+
76 – 80	B
71 – 75	B-
66 – 70	C+
61 - 65	C
56 – 60	C-
51 – 55	D+
46 – 50	D
0 – 45	D-