TO: The Faculty of the College of Engineering

FROM: The Faculty of the School of Mechanical Engineering

RE: ME 23900 Introduction to Data Science for Mechanical Engineers – Permanent Course Number

The Faculty of the School of Mechanical Engineering has approved the following new course. This action is now submitted to the Engineering Faculty with a recommendation for approval.

ME 23900 Introduction to Data Science for Mechanical Engineers
Credits: 1 credit
Offered: Spring, Summer, and Fall
Prerequisite: MA 261 (Multi-variate calculus)

Course Description: Introduction to data science for sophomore mechanical engineers with no prior knowledge of the topic; Overview of Python basics using Jupyter notebooks; review of probability theory as the language of uncertainty, Monte Carlo sampling for uncertainty propagation; basics of supervised (Bayesian generalized linear regression, logistic regression, deep neural networks, convolutional neural networks), and unsupervised learning (k-means clustering, principal component analysis, Gaussian mixtures).

Learning Outcomes:

- Represent uncertainty in parameters in engineering or scientific models using probability theory
- Propagate uncertainty through physical models to quantify the induced uncertainty in quantities of interest
- Solve basic supervised learning tasks, such as regression and classification
- Solve basic unsupervised learning tasks, such as clustering, dimensionality reduction, and density estimation
- Develop and apply various Python coding skills
- Load and visualize data sets in Jupyter notebooks
- Visualize uncertainty in Jupyter notebooks
- Recognize basic Python software (e.g., Pandas, numpy, scipy, scikit-learn) and advanced Python software (e.g., pytorch, pyrho, Tensorflow) commonly used in data analytics

History: This course has previously been offered as a ME 297 course titled "Introduction to Data Science for Mechanical Engineers" three times since Fall 2020. The student enrolment in the course was as follows:

- Fall 2020 with 14 students
- Spring 2021 with 17 students
- Fall 2021 with 8 students

Background: With the transition from MA 272 (4 cr) + MA303 (3 cr) sequence to MA 265 (3 cr) + MA 266 (3 cr) sequence, the required math credits for BSME will be reduced by one credit. The School of Mechanical Engineering has voted to make ME 239 a core requirement for BSME. The current EFD is for permanent course number only. A separate EFD will be submitted for the updated BSME plan of study.

Jitesh Panchal
Associate Head for Undergraduate Programs
Professor of Mechanical Engineering

ME 297: Introduction to Data Science for Mechanical Engineers

Instructor

- Ilias Bilionis, *Associate Professor of Mechanical Engineering, Purdue University*

Audience

Mechanical engineering sophomores

Course Description

This course provides an introduction to data science for sophomore mechanical engineers with no prior knowledge on the topic. The course starts with a brief overview of Python basics making use of Jupyter notebooks (the most popular computational platform in modern data science). Then, it covers and extensive review of probability theory as the language of uncertainty, discusses Monte Carlo sampling for uncertainty propagation, covers the basics of supervised (Bayesian generalized linear regression, logistic regression, deep neural networks, convolutional neural networks), and unsupervised learning (k-means clustering, principal component analysis, Gaussian mixtures). Throughout the course, the instructor follows a probabilistic perspective that highlights the first principles behind the presented methods with the ultimate goal of teaching the student how to create and fit their own models.

Course Learning Outcomes

After completing this course, you will be able to:

- Represent uncertainty in parameters in engineering or scientific models using probability theory
- Propagate uncertainty through physical models to quantify the induced uncertainty in quantities of interest
- Solve basic supervised learning tasks, such as: regression, and classification
- Solve basic unsupervised learning tasks, such as: clustering, dimensionality reduction, and density estimation
- Develop and apply various Python coding skills
- Load and visualize data sets in Jupyter notebooks
- Visualize uncertainty in Jupyter notebooks
- Recognize basic Python software (e.g., Pandas, numpy, scipy, scikit-learn) and advanced Python software (e.g., pytorch, pyrho, Tensorflow) commonly used in data analytics

Prerequisites

- Working knowledge of multivariate calculus and basic linear algebra
- Some programming experience, e.g., Matlab. Basic Python knowledge would be helpful but not required
- Knowledge of probability and numerical methods for engineering would be helpful, but not required

Required Tools

Jupyter Notebook

Jupyter notebooks are interactive documents that can simultaneously contain text, mathematics, images, and executable code. The executable code can be in many programming languages (e.g., R, Matlab), but we are only going to use Python in this course. The course uses Jupyter notebooks for the following content: Reading Activities, Hands-on Activities, and Homework Assignments. The rationale behind this choice is that it allows the student to focus on the mathematical methods rather than the programming and it ensures the reproducibility of the course content. Of course, understanding the code in Jupyter notebooks does require knowledge of Python, albeit it does not require knowing how to structure and call Python code from the command line. Jupyter notebooks can be run either on the students' personal computers (instructions vary with operating system and can be found here) or in several cloud computing resources. The recommended method for this class is to use Google Colab which is available free of charge and requires only a standard Google account. The activity links included in the course will take you automatically to a copy of the latest version of the corresponding Jupyter notebook which you can then save and edit on your Google Drive. If you do not want to create a Google account, then it is also possible to run the notebooks at Purdue's Jupyter Hub using your Purdue Career Account. This method requires more expertise and you may encounter several issues. When this occurs, please post a question on Piazza and the TA will help you out.

Access to the Jupyter Notebooks Repository

As stated earlier, the recommended method for using the Jupyter notebooks of this class is to use Google Colab. The links to all the activities will take you directly to a Google Colab copy of the Jupyter notebook. If you want to use any alternative method (e.g., your personal computers Purdue's Jupyter Hub, or anything else), you will need access to the Jupyter Notebook repository for the class. Select this link for a Git version control repository.

(If you have no idea what Git is, this is a good tutorial.) All the activities are inside the "activities" folder and they are named using the following convention: "activity_type_<lecture_number>.<activity_type_count>.ipynb." For example, "hands-on-06.1.ipynb" is the first hands-on activity of Lecture 6.

Grading

This course will be graded based on the following criteria:

Assessment Type	Description	% of Final Grade
Homework	There will be eight (8) homework assignments. The homework assignments will be both theoretical (e.g., prove this, derive that) and computational (e.g., use this data to fit that model, create and fit a model for this situation). The assignments will be in the form of a Jupyter notebook with empty space reserved for your writing or coding. If you wish, you can do the writing by hand (instead of the latex required by Jupyter notebooks), scan it and submit a single PDF. Submissions should be made through Gradescope.	80%
Quizzes	There is a total of thirty eight (38) quiz assignments in the class. The difficulty varies, but they are all worth the same. The quizzes always follow a video (and usually the answer is in the video). You only get one chance to answer the quiz questions. So, do it carefully. You have until Sunday 11:30pm each week to submit the quizzes for every week. Discuss the quiz answers in Piazza or bring them to the office hours.	20%

Grading Scale

Your course grade will be based on the following grading scale: A+ 98 – 100; A 88 – 98; A- 85 – 88; B+ 80 – 85; B 73 – 80; B- 70 – 73; C+ 67 – 70; C 62 – 67; C- 60 – 62; D+ 57 – 60; D 52 – 57; D- 50 – 52; F 0 – 50. The instructor reserves the right to change with grading scale in a way that favors students.

Course Schedule

Section	Lectures	Assignments
Introduction	0. What is data science and why is it important for mechanical engineers (1/19) **Note:** There is no physical lecture, because 1/19 is MLK day, but there reading assignment for the week. Material will be made accessible through Brightspace.	NONE
Python Basics	1. Basic data types, operators, functions, conditionals, and loops (1/19) **Note:** There is no physical lecture, because 1/19 is MLK day, but there reading assignment for the week. Material will be made accessible through Brightspace. 2. Advanced types, numerical arrays, and plotting (1/25) **Note:** This is the first in-person lecture.	• **Homework 1** ○ Available in the course 1/19 ○ Due Date: **1/31 at 11:59 PM ET** *(2/1; 3:59 UTC)*
Review of Probability Theory	3. Basics of Probability Theory (2/1) 4. Discrete Random Variables (2/8) 5. Continuous Random Variables (2/15) 6. Collections of Random Variables (2/22)	• **Homework 2** ○ Available in the course: 2/1 ○ Due Date: **2/14 at 11:59 PM ET** *(2/15; 3:59 UTC)* • **Homework 3** ○ Available in the course: 2/15 ○ Due Date: **2/28 at 11:59 PM ET** *(3/1; 3:59 UTC)*
Uncertainty Propagation	7. The Monte Carlo Method for Estimating Expectations (3/1)	• **Homework 4** ○ Available in the course: 3/1 ○ Due Date: **3/7 at 11:59 PM ET** *(3/8; 3:59 UTC)*
Principles of Bayesian Inference	8. Analytical Examples of Bayesian Inference (3/8)	• **Homework 5** ○ Available in the course: 3/8 ○ Due Date: **3/14 at 11:59 PM ET** *(3/15; 3:59 UTC)*
Supervised Learning: Linear Regression and Logistic Regression	9. Linear Regression Via Least Squares (3/15) 10. Bayesian Linear Regression (3/22) 11. Classification (3/29)	• **Homework 6** ○ Available in the course: 3/15 ○ Due Date: **4/4 at 11:59 PM ET** *(4/5; 3:59 UTC)*
Unsupervised Learning	12. Clustering and Density Estimation (4/5) 13. Dimensionality Reduction (4/12)	• **Homework 7** ○ Available in the course: 4/5 ○ Due Date: **4/18 at 11:59 PM ET** *(4/18; 3:59 UTC)*
Neural Networks	14. Deep Neural Networks (4/19) 15. Deep Neural Networks Continued (4/26)	• **Homework 8** ○ Available in the course: 4/19 ○ Due Date: **5/1 at 11:59 PM ET** *(5/2; 3:59 UTC)*

Course Help

To get help with course content, comment in the Piazza discussion forums. Each homework assignment and lecture has its own corresponding discussion in the Piazza course site. By commenting in these discussion forums, the course team will be able to respond to your question more quickly.

Discussion Guidelines

Please follow the Discussion Guidelines when contributing to discussions in this course. Here are a few of the key points you should remember:

- Do not use offensive language. Present ideas appropriately.
- Be cautious in using Internet language. For example, do not capitalize all letters since this suggests shouting.
- Avoid using vernacular or slang language. This could possibly lead to misinterpretation.
- Do not hesitate to ask for feedback.
- Be concise and to the point.
- Think and edit before you push the “Send” button.

Accessibility Information

Purdue University strives to make learning experiences as accessible as possible. If you anticipate or experience physical or academic barriers based on disability, you are welcome to let me know so that we can discuss options. You are also encouraged to contact the Disability Resource Center at: drc@purdue.edu or by phone: 765-494-1247.

The **Office of Institutional Equity**, which is responsible for ensuring Americans with Disability Act compliance, can be contacted with any accessibility concerns at:

Phone: (765) 494-7253
Email: equity@purdue.edu
TTY: (765) 496-1343
Website

- Purdue's Disability Resource Center Website
- Purdue's Web Accessibility Policy
- Purdue Equal Access Frequently Asked Questions (FAQs)
- Google Accessibility
- Jupyter Notebook a11y toolbar (from Microsoft research)

Academic Integrity

Academic integrity is one of the highest values that Purdue University holds. Individuals are encouraged to alert university officials to potential breaches of this value by either emailing or by calling 765-494-8778. While information may be submitted anonymously, the more information that is submitted provides the greatest opportunity for the university to investigate the concern.

The Purdue Honor Pledge

"As a boilermaker pursuing academic excellence, I pledge to be honest and true in all that I do. Accountable together - we are Purdue"

Nondiscrimination Statement

Purdue University is committed to maintaining a community which recognizes and values the inherent worth and dignity of every person; fosters tolerance, sensitivity, understanding, and mutual respect among its members; and encourages each individual to strive to reach his or her own potential. In pursuit of its goal of academic excellence, the University seeks to develop and nurture diversity. The University believes that diversity among its many members strengthens the institution, stimulates creativity, promotes the exchange of ideas, and enriches campus life. Link to Purdue's nondiscrimination policy statement.