ISyE 4602 Foundations of Modern Data Science

Syllabus

Course Information

Required/Elective: Concentration Depth for Analytics and Data Science, and AI/OR for Decisions

Credit: 3-0-3

Prepared by: Prof. Ashwin Pananjady, Fall 2026

Prerequisites: ISyE 3030 Basic Statistical Methods or ISyE 3770 Statistics and Applications; ISyE 3133 Engineering Optimization; CS 1301; or similar coursework

Catalog Description

This advanced undergraduate course introduces the mathematical principles behind probabilistic modeling, optimization-based model-fitting, and statistical inference, which form the bedrock of ML and AI systems.

Text

No formal textbook will be used.

References

Lecture notes and supplementary materials will be available.

Software

Python (NumPy, pandas, scikit-learn) will be used for assignments and projects. Use of Jupyter or Colab is encouraged.

Course Description

This course develops the mathematical foundations of data science, introducing students to the life cycle of data-driven decision-making based on principles of probabilistic modeling, optimization-based model-fitting, and statistical inference. Topics include frequentist and Bayesian paradigms of data science, low-dimensional parameter estimation and confidence intervals, inference and uncertainty quantification in linear and nonlinear regression, high-dimensional models and fundamental limits of inference, high-dimensional stochastic optimization problems, and multiple hypothesis testing. Time permitting, the course will also cover some basic causal inference and robust data analysis algorithms. The focus will be on looking under the hood of modern data science methodology. While there will be some exposure to implementing these methods on real data, the primary focus will be on understanding these methods and developing judgment for them.

Topical Outline

Part I: Prediction

- Basic statistical principles: generative modeling, Bayesian and frequentist thinking, maximum likelihood, and MAP. Rigorous confidence in decision-making and the assumptions on which it relies.

- Linear regression, including reviews of multivariable calculus and linear algebra.
- Nonlinear regression, in particular polynomial regression.
- Regularization, bias-variance trade-off, and validation on holdout data.
- Predicting probabilities: classification and logistic regression.
- Nonparametric methods: nearest neighbors.
- Classification with neural networks.
- Fundamental understanding of optimization algorithms for prediction problems.
- Quantifying model credibility through statistical analysis.

Part II: Inference and Confident Decision-Making

- Cross-validation and resampling.
- Bootstrapping.
- Binary decisions, p-values, and confidence intervals.
- A/B testing with parametric and nonparametric methods.
- Multiple testing.
- Sequential testing using multi-armed bandits.
- Observational data and causal inference.

Bonus Topics in Data Science

These topics will be covered only if there is time and interest:

- Robustness.
- Sequential decision-making via reinforcement learning.

Course Learning Outcomes

Upon successful completion of the course, you will have learned:

- How to recognize and justify a reasonable data generating process for a problem at hand.
- Design principles for data analysis based on formulating and solving optimization problems.
- How to think carefully and critically about modeling assumptions under which parametric and nonparametric methods are expected to work, and to understand the downstream goals of prediction and decision-making.
- To have a qualitative and quantitative understanding of which method is more suitable given the various constraints, including data, computation, and flexibility, at hand.

Mapping from Course Outcomes to ABET Student Learning Outcomes

Table 1 shows the relationship between course outcomes and ABET student learning outcomes. The first column lists the course learning outcomes. The next seven columns list the abbreviated ABET student learning outcomes. The entry of H in a row with a specific course learning outcome and column with a specific ABET student learning outcome indicates that the importance of attaining this outcome is high and will be assessed.

Table 1. Mapping from course outcomes to ABET student learning outcomes

Course outcome / Program outcomes	1. Identify, formulate, and solve engineering problems by applying engineering, science, and mathematics	2. Produce solutions that consider public health, safety, welfare, global, cultural, social, environmental, and economic factors	3. Communicate with a range of audiences	4. Recognize ethical and professional responsibilities and make informed judgments	5. Function effectively on a team, provide leadership, create a collaborative and inclusive environment, plan tasks, and meet objectives	6. Develop and conduct experiments, analyze and interpret data, and use engineering judgment to draw conclusions	7. Acquire and apply new knowledge using appropriate learning strategies
1. Recognize and justify a reasonable data generating process	H					H	H
2. Design principles for data analysis based on formulating and solving optimization problems						H	H
3. Think critically about modeling assumptions and understand the downstream goals of prediction and decision-making	H	H		H	H		
4. Understand which method is more suitable given the various constraints, including data, computation, and flexibility		H			H		

Evaluation of Important Outcomes

H will be assessed in the final exam or final project.