

ISYE 4602 FOUNDATIONS OF MODERN DATA SCIENCE

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

Credit: 3-0-3

Prepared Prof. Ashwin Pananjady, Fall 2026

Prerequisite(s): ISYE 3030 Basic Statistical Methods or ISYE 3770 Statistics and Applications, 3133 ENGINEERING OPTIMIZATION and CS 1301 and ISYE 3133 Engineering Optimization or Similar

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/projects. Use of Jupyter/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, we will also cover some basic causal inference and robust data analysis algorithms. Our 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, our primary focus will be on understanding these methods and developing judgement 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 (reviews of multivariable calculus and linear algebra)
- Nonlinear (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 edibility 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 (will be covered only if there is time/interest)

- Robustness
- Sequential decision-making via reinforcement learning

Course Learning Outcomes

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

1. How to recognize and justify a reasonable data generating process for a problem at hand.
2. Design principles for data analysis based on formulating and solving optimization problems.
3. 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.
4. To have a qualitative and quantitative understanding of which method is more suitable given the various constraints (data/computation/flexibility) at hand.

The Mapping from the Course Outcomes to ABET Student Learning Outcomes

The table below shows the relationship between course outcomes and ABET student learning outcomes. The table has 8 columns. The first column lists the course learning outcomes listed above. The next 7 columns list the abbreviated ABET Student Learning Outcomes. The entry of H in a row with specific course learning outcome and column with specific ABET Student Learning Outcomes indicate the importance of attaining this outcome is high and will be assessed.

Course outcome \ Program Outcomes	1. identify, formulate solve engg prob by engg, sci & Math	2. produce solutions consider public health, safety, welfare, global, cultural, social, environ & economic	3 communicate with a range of audience	4 recognize ethical & professional responsibilities, make informed judgement consider resolutions in global, economic, environ and societal context.	5. effective on a team provide leadership, collaborative and inclusive environ, plan tasks & meet objectives	6. develop and conduct experiment, analyze and interpret data & use engineering judgement 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 to understand the downstream goals of prediction and decision-making	H	H		H	H		
4. Understand which method is more suitable given the various constraints (data/computation/flexibility) at hand		H			H		

Evaluation of the important outcomes

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

