COMP 585 **Spring 2025**

Probabilistic Toolkit for Learning and Computing *Instructor: Maryam Aliakbarpour*

Project Guidelines

The goal of the course project is to provide an opportunity for you to get involved with a current topic of research in learning theory or theoretical computer science, focusing on the essential use of randomness and probabilistic arguments. You may choose between two types of projects: a survey or original research. For a survey project, you need to write a survey based on **three** relevant papers. For original research, you need to strive to develop a new model, algorithm, or theorem, or you may focus on experimental results. For further information, see the appropriate sections of this document.

We have the following four assignments regarding the project throughout the semester. For deadlines, see Table 1.

- **Proposal:** For this assignment, write a one-page project proposal outlining your project plans. Within this report, include the following:
  - Problem definition: Define the topic or the problem you wish to work on.
  - Motivation: Explain why this topic is significant both to you and to the broader research community. Highlight its relevance to the course topic.
  - Literature review: Conduct a preliminary review of the existing results. Identify key papers that are related to your project.
  - Your plan: Describe your proposed project and its scope in detail. In addition to your final (potentially ambitious) goal, outline achievable steps that can be accomplished within one to two weeks. Identify the initial steps that you will take or have already taken before the mid-point evaluation.
- **Mid-point evaluation:** This is a milestone where you aim to finish $\sim 40\%$ of your project. You need to produce a three-page report for this assignment. Think of this report as a mini final report. Explain the partial results you may have. What has worked so far and what did not quite work.
- **Final report:** Write an six-page final report for the project. See the relevant sections for more details on what a good project report consists of.
- **Project presentation:** We will have project presentations at the end of the semester.

Assigment	**Deadline**
Proposal	03/13/2025
Mid-point evaluation	04/03/2025
Project presentation	Last two lectures of class: 04/22/2025 and 04/24/2025
Final report	04/24/2025

Table 1: Assignments and deadlines

Policies

Novelty: What you do for the class project should be novel and original; you may not reuse materials that you have already submitted for publication or coursework in other classes.

Late submission: You may lose your late days for your submission, but keep in mind that these are also shared with your deadlines for the problem sets. Beyond that, you will lose 10% of your grade per late day. For the final report, submit it no later than 04/27/2025.

Format: Please typset your reports for these deliveries in Latex and upload them on Canvas by the indicated deadlines.

Group project: You may pair up with another member of the class. However, the expectation for the group projects will be higher accordingly.

Rice Honor Code: You are expected to adhere to the Rice Honor Code. You are encouraged to collaborate and find resources online. However, all the material to be graded is expected to be original unless properly recognized and cited. This policy includes the use of large language models (such as chat-gpt). It is permissible to apply such software for spell/grammar checks to your original text. However, these tools are prohibited for generating content that is not deemed to be yours, including rephrasing others' work and producing summaries.

Surveys

Select a minimum of **three related research papers**, read them carefully, and write a survey. You may choose your own set of papers as long as they are relevant to the topic of the class. A good survey consists of the following:

- **Motivation:** Describe the significant real-world problem that the papers you have chosen aim to address and explain its importance.

- **Literature review:** In addition to the three papers you have selected, compile a comprehensive list of papers related to the topic and explain their differences.
- **Problem definition:** For each paper, provide a clear, formal statement of the problem.
- **Technical overview of the results:** For each paper, provide an overview of their contributions.
- **Comparisons and connections:** This survey should go beyond summarizing papers – it must present the connections between them. For example, explain how one paper builds on another or what novel techniques one paper used that allowed them to overcome barriers others could not.

Original research

For the class project, you can work on a novel research problem and either solve it or take steps toward a solution. This could be an extension of your current research or involve proposing experimental verification of an existing result. If you have a problem in mind to propose, that's fantastic. If you need assistance in finding one, I would be happy to share a few problems that I am interested in working on as well.

A good final project report for original research consists of the following:

- **Motivation:** Describe what significant real-world problem this research project aims to address and explain its importance.
- **Problem definition:** Provide a clear, formal statement of the problem.
- **Literature review:** Compile a comprehensive list of papers related to your work and explain the advantages and the differences in models or assumptions compared to your work.
- **Technical overview of your results:** Think of this section as what a reviewer would read instead of reading all your proofs. Provide a high-level explanation of your techniques and highlight the significance of your work.
- **Your contributions:** Describe your results in detail here. You may also discuss the techniques you have worked on that did not yield successful results. Try to identify the cases where your solution does or does not work.

Suggested Topics

You may choose from a broad set of topics in statistical learning theory and theoretical computer science. Below are a few interesting topics that you might want to explore. You may also come to the office hour to chat about your project.

1. **Augmented Algorithms**

   Augmented algorithms are those that receive predictions or external inputs (such as machine learning predictions, real-time data, or human feedback) to enhance their decision-making or improve performance. This framework bridges the gap between theory and practice, as potentially unreliable predictions (e.g., those generated by an ML model) can be leveraged by a rigorous algorithm to produce results with provable guarantees. Additionally, when the predictions are accurate, they can significantly improve the algorithm's efficiency.

2. **Meta-Learning**

   Meta-learning, or "learning to learn," is a framework in which models are trained to adapt quickly to new tasks with minimal data. This approach is particularly useful when working with a heterogeneous set of users learning a series of related but not necessarily identical tasks.

3. **Calibration**

   Calibration in machine learning refers to the alignment of predicted probabilities with actual outcome frequencies. Well-calibrated models provide reliable confidence estimates, which are essential for high-stakes applications such as medical diagnosis and autonomous driving. A well-calibrated model is expected to provide accurate uncertainty estimates, leading to better decision-making and risk assessment.

4. **Private Data Analysis**

   Private data analysis focuses on extracting useful insights from sensitive data while preserving privacy. Differential privacy (DP) is a key technique that ensures individual-level data cannot be inferred from aggregate results. Advancements in this field include robust data-sharing frameworks that enable privacy-preserving machine learning without compromising data utility.

5. **Machine Unlearning and Its Connection to Differential Privacy**

   Machine unlearning aims to selectively remove learned information from a model, often to comply with data deletion requests or regulatory requirements such as the GDPR. Connections to differential privacy arise because both involve controlling a model's sensitivity to individual data points. The main goal is to maintain the trained model efficiently while ensuring that, within a given deletion budget, the model remains nearly identical to one that never used the deleted data point, in a provable manner.