Prof. Audrey Sedal

MECH 501 – Probability, Statistics and Machine Learning in Mechanical Engineering

Fall 2023

Assignment 2 Details

Assignment 2 is an individual project. In this project, there are four key tasks.

**Task 1:** Find a problem from the domain of mechanical engineering.

**Task 2:** Find a data set, simulator, or simulation framework related to that problem.

**Task 3:** Use the data set/simulator with techniques covered in the course to attempt to solve the problem.

**Task 4:** Communicate your results in a Jupyter notebook and a short presentation.

Every student will make a 10-min appointment with Prof. Sedal to discuss and confirm their choice of problem. Students will provide a pitch explaining the problem, the data set or simulator, and the planned technique. They will justify the scope and novelty based on the details given below. Prof. Sedal will give feedback on it. Appointment availability will be given via a myCourses announcement. Following the appointment, students will complete the project before the due date.

Detailed Instructions

* **Task 1:** Find a problem from mechanical engineering.

1. The problem must pertain to mechanical engineering (ME).
   * 1. Example of ME problem: “Can convolutional neural networks (CNNs) be used to guess the missing data in a sequence of fluid flow measurements?”.
     2. Example of non-ME problem: “Can CNNs be used to guess whether there is a bird in a photo?”
2. The problem should be framed as “Can [X] be used to do [Y]?” where X is a probability, statistics, or ML technique and Y is prediction or estimation-related task.
3. The problem should be narrow in scope and tractable for you.
4. Resources to find problem ideas are given in Appendix A. If you are currently doing research, you may use a problem that intersects with your research.

* **Task 2:** Find a data set or simulator related to the problem.
  1. The data set or simulator must be published in an academic venue (conference or journal) and freely available.
  2. You may not collect new data for this project. If you use data that you previously collected, it must be published (as above) or be included in a pre-print.
  3. Similarly, you may not create a new simulator or simulation. If you use a simulation you have previously created, it must be published or included in a pre-print.
  4. Similarly to above, resources are given in Appendix A.
* **Task 3:** Use the data/simulation with techniques covered in the course to attempt to solve the problem.
  1. You can use any technique covered in-depth in the course (any point of the course), a technique mentioned in the course, or a technique that extends from the course concepts.
  2. You may not repeat problem-technique pairs that are famous, that exist as textbook problems or examples, or that repeat coursework or research which you have previously completed. However, you may propose an extension of your own prior work.
  3. You are advised to try a few varied implementations, where applicable, and either justify it if you only do one or compare a few implementations.
  4. Your pitch to Prof. Sedal should have details about how you will implement the technique and your own experience (if you have any) in the domain.
* **Task 4:** Communicate the results in a Jupyter notebook and a short presentation.
  1. The grader should be able to re-generate quantitative results from your work (e.g., re-run your technique); the notebook should be submitted alongside any supplementary material needed to make it work. Error/failure of the notebook to run due to missing references can result in points not being awarded.
  2. Your notebook should:
     1. Motivate the problem (in words, i.e. markdown)
     2. Explain the technique (in words)
     3. Provide readable, clear code that implements the technique correctly
     4. Draw conclusions about the technique’s performance in the problem context.
  3. It is admissible if the technique does not work as hypothesized. However, the technique still needs to be implemented thoroughly.
     1. Example of “good” failure. Problem: Can a Particle Filter (PF) predict the motion of this robot better than an Extended Kalman Filter (EKF)? Method: The technique was tested with a large enough data set and a few different distributions of the PF. Conclusion: In this robotics data set, the EKF and PF had comparable error. Therefore, the PF did not outperform the EKF.
     2. Example of “bad” failure: Problem: Can a CNN predict the missing fluid data? Method: A CNN was trained with an un-motivated choice of hyperparameters and training fails. No fine-tuning was attempted. Conclusion: The CNN did not predict the missing data, because it was never trained, because no hyperparameter tuning was done.
  4. Your short presentation should summarize the problem, technique and conclusion.

Assessment

**Baseline requirements.** These are needed to earn any points at all:

* Problem pertains to ME
* Problem is correctly scoped (per above) and tractable
* All tasks are complete

**Assessment Criteria.** A project can earn 100% if it fits all of the below characteristics.

*Motivation (20%)* - The problem is described clearly so that a technical-non expert can understand it. Its importance in engineering (industry or science) is well-explained, and appropriate sources are given. The choice of technique for the problem is well-justified.

*Method (50%)* – The technique is clearly explained in words in the presentation. The data set or simulation matches with the problem. Diagrams or algorithm tables are given where appropriate. Auxiliary choices (e.g., parameters, hyperparameter tuning techniques) are justified matching with existing norms for that technique. E.g., where relevant, ablations are performed. All code runs in the notebook without bugs, and results are reproducible; stochastic/randomness in the code does not create wildly different conclusions each run. Code is well-formatted, easy to read, and commented.

*Writing Quality (5%)* – Markdown sections of the notebook are clear and easy to read. Appropriate sections are used. References are given, student’s name is on the notebook, etc.

*Conclusions and Discussion (25%)* – A tentative answer to the problem is written down (in the report) and justified (in the spoken presentation). Data visualizations or other data communication techniques are clear and support the written/presented discussion points. Limitations of the technique are well-clarified in the case of success or failure. For successes, possible subtle reasons why it worked are given.

Time

This project is expected to take up to 30 hours. Estimates: 5-10 hours toward finding the problem and data set, 10-15 hours toward programming, and 5 hours toward creating a high-quality notebook and presentation. However: it is impossible to predict the exact amount due unforeseen circumstances, e.g.: difficulties installing needed libraries on the machine one has access to (adds time), already having a topic in mind (reduces time).

Plagiarism and Academic Conduct

Please see all academic conduct notes as given on the course outline.