

MTH 4320/5320 Project 1

Proposal Deadline: Sept 19

Project Report Deadline: Oct 13

Presentations: Oct 14-15 (details TBA)

Project Description

You will find a data source or gather data to use a **fully-connected feedforward neural network** for a well-defined regression or classification problem (or a mixed problem).

The project will require preprocessing the data, tuning hyperparameters, and solving the problem as well as possible. You will need to test different normalization methods, network architectures, activation functions, loss functions, regularization methods, initialization methods, and/or modifications to the learning method.

You may work in teams or individually. Project expectations will be proportional to the number of team members, so you should only work as a team if more work or diverse skill sets are necessary to the problem.

Proposal

For each project, you must submit a proposal (max 1 page) including the following:

1. The **goals** of your study.
2. Your **intended data sources**. There must be thousands of datapoints, preferably tens of thousands.
 - (a) You can use existing labeled data.
 - (b) You can use existing data and clean/preprocess it.
 - (c) You can generate your own data (data from sensors or simulations are good ideas since it can be automated).
3. **Team:** If you want to work in a team, list who will be on the team, where each person plans to focus in the project.

The purpose of the proposal is to design a feasible study at an appropriate level. I will evaluate them and give feedback to ensure the level of difficulty is appropriate given the team size.

Proposal Grading

The proposal is worth 20 points. The criteria are:

- 10 points: The goals of the regression and/or classification problem is described in detail.
- 5 points: Appropriate data has been located for use or you have a plan to gather the data.
- 5 points: You have a reasonable plan for cleaning/preprocessing the data.

Project

Each student or group will attack your well-defined regression or classification problem with your real-world data set.

Each project is different and different projects will require more work on some parts than others, but all should have four main parts:

1. **Data preparation:** Create or clean the data (if needed), read the data with code, and preprocess the data.
2. **Benchmarking:** Use a simple model (for example, linear regression or logistic regression) to find some benchmarks for performance.
3. **Training and Tuning your NN:** Set up a fully-connected feedforward neural network and tune the hyperparameters to solve your problem as well as possible.
4. **Presenting your work:** Write a report. There are no rules on length, just thoroughly describe your work. Some guidelines:
 - Include an introduction and discussion of your problem and dataset.
 - Include your code, written descriptions of your code, outputs of your code
 - Include results of the hyperparameter experiments you run (not necessarily every one, but show me you tested things that worked or didn't work).
 - Keep in mind preprocessing, normalization methods, initialization, activation functions, loss functions, regularization methods, learning rates or schedules, net architectures and any other decisions you make for your model are all hyperparameters.
 - Explain the reasoning for your progression in adjusting the hyperparameters.
 - Include performance metrics or graphics for the experiments.
 - Present your work to the instructor/class (details TBA)

Project Grading

The project is worth 200 points. (Students within teams will receive the same grade.)

- 20 points: Proposal
- 20 points: Problem description in report (What are you trying to do? What is your data? What shape is it?)
- 20-60 points: Data preparation (Include any code, describe your process.)
- 20 points: Benchmarking (Which method(s) are you using? What is the performance?)
- 40-80 points: Training and tuning your network (include discussion of hyperparameter decisions and results of experiments)
- 20 points: Conclusions, descriptive diagrams, etc., in the report
- 20 points: Present your work (5-10 minutes)

Note that the amount of points dedicated to some categories depends on the nature of the project as determined by the instructor because some projects require more work on the data or in hyperparameter tuning than others.