

Group Research Assignment: Dataset Selection

Theory of Machine Learning



Begin with choosing a team name based on X-men characters. I am now and forever Professor Z (like Professor X from X-Men), and we will now refer to 625.742 as the X-Mansion (maybe the JH-Mansion to make sure no copyright laws are violated ☺). Sense so many people see machine learning as a special magical ability, I will refer to my students as my mutant students (students with special gifts). Welcome to the Mansion, you are here because of your exceptional gifts, and I am here to help you master them! I know it's cheesy, appease me please! ☺

Choose a dataset that is of interest to your group. Do not pick a dataset that is really easy or well understood. I challenge you to pick on that is less understood and research about the dataset. This will make you learn more about novel data and will guarantee that you get the most from the course. Your assignment is to research your dataset. Give background information about the dataset, find all or make reasonable assumptions about the dataset based on background information or from experts in the field, find out what predictive power is hypothesized about the dataset, tell why the dataset is important. This will be presented, using a tool of your choice ([Zoom](#), [YouTube](#), [Vimeo](#), etc.), in a 20-minute presentation by your group at the midpoint of the course (Module 7).

The goal of the project is to not get a dataset that has a guaranteed answer too. If during the final presentation, substantial work has been done but no final results have been achieved, the project is considered a success. If the group is having a hard time with this, please see me (Professor Z) during office hours or email me and we can set up an appointment with your group.

By Day 3 of Module 7, your presentation should be complete and shared in Discussions. Each student is expected to review **all** other teams' presentations using the provided "Rubric For Group Oral Presentations" and, **by Day 5 of Module 7**, ask **at least one question to at least one group**; however, more questions are encouraged throughout the week. **By Day 7 of Module 7**, every group should respond to the questions they have been asked.

