

Project Assignment : Study Plan

Due date: October 18, 2021

Please submit a one-page summary of your proposed research question and study to Gradescope.

The instructor will meet with each group to discuss study plans during the lecture on **October 25**.

The instructor will send you a schedule the day before. There will be probably only about 15 minutes so please make sure that your study plan is clear and precise. You may also include questions that you would like to discuss at the meeting at the end of the document.

To find interesting datasets for your project, you can check:

- **Kaggle**: <https://www.kaggle.com/competitions>
- **UCI: Machine Learning Repository**: <https://archive.ics.uci.edu/ml/datasets.php>
- **Awesome public datasets**: <https://github.com/awesomedata/awesome-public-datasets>

Some advice for selecting the project:

- **Be selective!** Don't choose a project that has nothing to do with machine learning. Don't investigate an algorithm that has a high chance of failing or being un-implementable. Don't attack a problem that is irrelevant, ill-defined or unsolvable. Spend most of your time doing machine learning and not related things such as pre-processing your data.

- **Be honest!** You are not being marked on how good the results are. It doesn't matter if your method is worse than the ones you compare to provided you implemented it properly. What matters is that you try something sensible and clearly describe the problem, your method, what you did, and what the results were.

Be modest! Don't pick a project that is way too hard. Usually, if you select the simplest thing you can think of to try, and do it carefully, it will take much longer than you think.

- **Be careful!** Don't do foolish things like test on your training data, set parameters by cheating, compare unfairly against other methods, include plots with unlabeled axes, use undefined symbols in equations, etc. Do sensible cross-checks like running your algorithms several times, leaving out small parts of your data, adding a few noisy points, etc. to make sure everything still works reasonably well. Make lots of pictures along the way.

- **Learn!** The point of the project is to give you a chance to "test drive" the process of doing machine learning. Consider this an opportunity to learn how to write code to run large experiments, make nice figures, layout readable equations, describe your work concisely to a smart but uninitiated reader, etc.

- **Have fun!** If you pick something you think is cool, that will make getting it to work less painful and writing up your results less boring.