Who is on the team?

Joe Sutker and Jason Lan

In a couple of paragraphs, describe the key ideas of your proposed project? What is your MVP? What are your stretch goals?

Our MVP will be a ground-up version of a commonly used machine learning algorithm with a poster explaining the main procedures involved in said algorithm. This poster will also contain a side-by-side comparison with the Scikit Learn version of the algorithm in order to show efficiency and accuracy. Another thing we need to do is to put the data from our poster onto a public website to help others easily understand the algorithm that we used. As a stretch goal, we're hoping to make our algorithm as efficient/accurate, or potentially more so, so that we can contribute it to the open source community.

To the best of your current knowledge, what datasets will you use for your project? Are there any obstacles you foresee in terms of getting access to the data?

We are planning on using a dataset from Kaggle. This would potentially be a dataset we've already used so that we can start in on the algorithm itself sooner, but we won't necessarily have already used this dataset. Since the data is from Kaggle, there should not be any obstacles to getting the data.

What are the most important new skills / techniques you will have to learn to be successful in this project? If you think some of these skills would be useful for us to cover in class, please indicate which ones.

The math involved in some algorithms and how each algorithm works (instead of just a black box) are very important things to know if you want to look deeper into the algorithms behind a lot of Data Science, instead of focusing on the visualization side.

Outline a rough timeline for the major milestones of your project. This will mainly be useful to refer back to as we move through the project.

Week 1: Choose dataset/algorithm, do background research on algorithm, begin learning inner workings of algorithm
Week 2: Finish learning inner workings of algorithm, begin implementation
Week 3: Implementation, continue learning inner workings
Week 4: Implementation, continue learning inner workings, comparison with scikit learn
Week 5: Documentation, implementation, continue learning inner workings
Week 6: Documentation, poster making, refining

What do you view as the biggest risks to you being successful on this project?

One of our biggest risks is the difficulty in understanding dense papers about our chosen algorithm.

Given each of your YOGAs, in what ways is this project well-aligned with these goals, and in what ways is it misaligned? If there are ways in which it is not well-aligned, please

provide a potential strategy for bringing the project and your learning goals into better alignment. There should be an individual section for each person on the team addressing the fit between the YOGA and the project topic.

Joe

This project is not necessarily well aligned with my initial first goal of making a neural network, but it could be if that's the direction we end up going in. It's much more well aligned with my revised first goal, since it basically *is* this project. This probably won't be well aligned with my initial second goal, since we probably won't be able to tie it into any of my hobbies, but it's ok, because the mathematics part of this project aligns really well with my interests, which is my revised second goal. We can definitely pair program on this project, which aligns really well with my third goal.

Jason

This project is well aligned with my goals as mine are using outside resources, thoroughly understanding an algorithm and gain more insights/ interpretations. My first goal can be reached by reading academic papers and watching online open courses throughout the project(which is what we are planning to do). My second goal is basically what we are trying to achieve in this project. And as for gaining insights, I'll constantly remind myself of extracting more thoughts and information from the knowledge we have at each stage of the project.