**PROJECT PROPOSAL FORM**

This document serves as an example. You can either fill out and submit, or use as a guideline for writing your own document.

The purpose of the proposal to demonstrate that you have put time into planning. Your final project can be different if you change your mind. The guidelines here are flexible, not hard constraints.

Student 1: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Student 2: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Student 3: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Student 4: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Date submitted: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Propose a title for your project.** If your project were written up as a research paper, what title would you give it? A good paper title will help each individual reader to know whether they should or should be interested in reading the paper. For example, the title [*Intriguing properties of neural networks*](https://arxiv.org/abs/1312.6199) (Szegedy *et al*. 2014) is a title that, although a little too vague, at least suggests that the nature of the work is an investigation, and that the focus was neural networks, and that the results are surprising. As another example, [*The fastest pedestrian detector in the West*](https://authors.library.caltech.edu/94253/) (Dollar *et al*. 2010) is a fun title indicating that the goal is “pedestrian detection” and that the nature of the contribution is “speed.”

[Type your answer here.]

**Describe the goal of your project.** What are you trying to achieve? What “main question” are you trying to answer, or at least to provide evidence for? Secondary goals are OK, but you should still have a clear “main goal” or “main question.” From your description, it should also be clear whether your project is about: making better predictions for some application? speeding up training and/or predictions? simply comparing predictive performance and/or speed of several methods? assessing or comparing interpretability? understanding failure modes or sensitivities of some methods? Etc.

[Type your answer here.]

**Describe the data you plan to use.** One of the hardest steps for a good machine learning project is to find data that is truly suitable for your goals. Finding good data not the most fun part, but it’s one of the most important—after all, for machine learning it is “garbage in, garbage out”. Here are some things you should ideally know:

* What are the ‘modalities’ that apply to the data? (images, video, speech, text, tabular, categorical, numerical, time series, experimental measurements, etc.)
* What does an input look like? (show an example if possible, like an image, or a sound wave, or some features, or at least try to describe)
* For an example input, what does the desired output look like?
* How many training and testing samples will there be? Can some be realistically trained on a laptop, or is something more powerful needed?
* Anything special about how the training and testing data should be split?
* Might the data need preprocessing before you can use it?

[Type your answer here. Be concise.]

**Describe how you will measure “success.”** You should explain how you will know whether you have achieved the goal(s) that you described earlier. What does “success” look like? What does “failure” look like? Keep in mind that your project can still succeed (in the sense of a good grade!) even if the experimental results are bad—what is important is that your experimental results are *conclusive*! A bad project is one in which you cannot even tell whether the goal was achieved or not.

[Type your answer here.]

**Describe how work will be divided.** It is very important for everyone to have a meaningful role in the project. If one person (the most experienced person) does all the programming or writing, then everyone else in the group loses this important chance to gain experience. For example, if there is no way to “happily divide” the work because two group members want to work on the same part, that is totally OK and no one should feel guilty for wanting that; both group members can do their own version of that part of the project, and then the final report can say “two group members each implemented did this part, and their results {matched, didn’t match}” When two people attempt and come to different conclusions, that is interesting and a chance for everyone to learn!

[Type your answer here. High-level description only, like “Angela will train the neural networks, and Seyyed will preprocess the data and train the SVM. Both will write the report.”]

**List the main Python packages you expect to use.** PyTorch? TensorFlow? Scikit-learn? Special packages for working with your data? (It is OK if this list is incomplete or changes for the final project.)

[Type your answer here.]