

Group Project Worksheet

This worksheet is intended to serve as a guide for your group projects. I am asking the class to self-organize into groups of 2-3 people and collectively work on a project that applies the concepts we have learned in the class. The projects will be due on the last day of the study period: May 8th. Every group should make a Github repository for the project and include its webaddress in Q1 below.

Broadly speaking, I envision the project falling into two classes:

- Implementing an existing algorithm/architecture from a ML paper using a package of your choice (Keras, TensorFlow, Pytorch). Interesting possibilities include writing your own RNN, Word2Vec, Neural Style Transfers, LSTMs, Transformers, a vision model. If the
- Using an existing model to answer a question about a *new dataset*. This could involve transfer learning or just using a pre-trained model. Regardless, the scientific question should be clear and in this case, I expect you to do lots of hyperparameter tuning to optimize performance.

However, other projects that apply the ideas and techniques we have learned these semesters are also acceptable.

In all cases, if the model or dataset is complex enough you may have to request resources from the SCC to implement the project. Please let me know ASAP if you need assistance with this. I imagine that in many cases the computing power of a laptop computer may not be sufficient for the project.

Please fill out the following worksheet and turn it in to me on Tuesday April 12.

1. Group Member Names/ Github Repository:

2. A brief summary of the project (maximum 250 word limit). Please complete this *after* answering the questions below.

3. Give a brief description of the datasets you will use in the project?

a. How big is the dataset?

b. What *data augmentation techniques* do you plan to use?

c. What is the feature space of the dataset?

d. How will you preprocess/normalize the data?

e. How will you split into training, test, and validation sets?

4. Briefly describe the model are you implementing to analyze this dataset?

a. Why are you choosing this model architecture and what advantages does it have for the dataset you are analyzing?

b. Please compare this architecture to other models that you could also use to analyze this dataset. What are the advantages? What are the disadvantages? (you can focus on computational resources, ease of coding, ease of training etc.)

5. What libraries/packages are you planning to use to implement the model?

a. Please explain why this is the right package/way of implementing the model

b. Are you planning to use GPUs to train the model? How will this change the kind of code you have to write?

7. Please explain how you will choose hyperparameters?

a. What are the hyperparameters you can tune in terms of model architecture (i.e. number of neurons etc)? How will you choose this hyperparameter?

b. What are the optimizers you plan to use including learning rates/learning rate schedule?

c. What forms of regularization do you plan to test? How will these be implemented in the code?

8. Please place the overall project in the context of the workflow in Figure 46 on p65.