Your task for the final project is to apply the deep learning tools and skills you have learned to a problem of your choosing that will have a positive social impact. Your deliverable for this project will be a blog post describing the problem, its impact, and how you've solved it.

Timeline

10/30 – 10/31: Finalize Groups You may work in groups of up to four on the final project. Please let me know your groups via this survey, and then create a repository for your group.

10/30 – 11/2: Brainstorm Topics You should generate *several* topic ideas that your group would be interested in working on. It is very likely that when you start searching for data, some ideas will turn out to be too ambitious and other will turn out to be trivial. Update your survey responses with your topic ideas.

11/2 – 11/6: Search for Data Sets For each of the topics you're considering, try to find multiple data sets you could use to tackle the problem. Your project is expected to incorporate data from multiple sources if possible. Update your survey responses with your proposed data sets.

11/7 – 11/13: Build and Test Preliminary Models Start small and try to get some sort of neural network to run on some subset of your data. Use these preliminary model tests to verify that your task is feasible and interesting, and finalize your topic choice on the survey. Think about how you will incorporate multiple data sets, and continue searching for additional data sources. Get a sense of the scale of the problem: the size of the data set, the size of the model, and the time (or other resources) required.

11/14 – 11/20: Run Experiments and Refine Models Run more thorough experiments for hyperparameter tuning, feature engineering, etc.; analyze your results and iterate; search the literature for related work on the problem; begin writing the background for your blog post.

11/21 – 12/1: Refine Experiments and Write Aim to finalize your models and their parameters so that your final experiments can focus on demonstrating performance. Take a step back and think about your report's narrative, and write a draft of the blog post.

12/2 – 12/8: Wrap Up & Revise Complete any pending experiments, revise and proofread the blog post, and submit by the end of finals.

Resources

Here are some resources you can use for identifying potential topics. This list is nowhere near exclusive; you should search for other resources too. If you have a topic you're interested in working on but need help finding ideas or data sets, feel free to contact Bryce for help.

- AI for social good workshops:
  - NeurIPS 2018
  - NeurIPS 2019
  - ICML 2019
  - ICLR 2019
  - note that each of these workshops have additional papers under the "accepted papers" tab
- Review articles covering many topics:
  - Tackling Climate Change with Machine Learning
  - AI for social good: unlocking the opportunity for positive impact
- Other papers
  - Combining satellite imagery and machine learning to predict poverty
  - Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings
- Blogs and websites
  - Data Science for Social Good

  - UN Global Pulse

Your group should brainstorm several potential project ideas. For each of the ideas you're considering, you should also identify possible data sets.

Model and Experiments

You are expected to build a model that combines multiple data sources. As with previous projects, you should try out several different model variations and compare them experimentally. You should establish *in advance* what criteria you will use to judge the success of your model. This means that when you choose your topic, you should consider what you will be measuring, and what results you hope to achieve according to those metrics. To avoid the temptation to move the goal posts, you are expected to submit your criteria on the survey by November 13th. You will not be graded based on whether you meet this target.

Write-Up

You will present your results in the form of a blog post written using a Jupyter notebook. The target audience for this writeup should be computer science majors who have not taken this class. Your goal should be to clearly explain the problem, what it means to solve it, how you used deep learning methods, how you measure success, and what results you achieved.

Your write-up should only include code that helps to tell the narrative of your project; other code can be imported from user-defined python modules. However, your writing should be sufficiently precise that a someone else in the class could replicate your results.

Here are some specific things to address in your report, in no particular order, and no matter the format of the report. This is not meant to be an exhaustive list.

- Where did you get your data? What preprocessing did you perform on the data? Describe your exploratory data analysis. Did you generate any plots or charts? Describe these, along with any relevant findings, in your report.
- What types of models did you try? Did you try any methods beyond the ones we've seen in class? How do the models you tried work? How did their performance compare to one another?

- What does success look like in this problem domain? How do you measure the performance of different models? What tradeoffs did you find between different models or parameters?
- What was your approach to building, training, and tuning models? How did you make hyperparameter choices? What parameters turned out to be most important? How did you address overfitting?

Your blog post should include citations for all of your data, and any resources you used in coming up with your models or experiments. The expected format for citations is a numbered list of web-links, along with a one-sentence description of the relevance of resource. These citations should be referenced elsewhere in the blog post, as appropriate.