Deep Learning Final Project

A group assignment

For this assignment, you should form groups of 2 to 3 people. Groups are to complete the assignment as a unit. Groups will be graded as a unit.

Overview

Your project is an opportunity for you to explore an interesting deep learning problem of your choice in the context of a real-world data set. This project can be whatever you want, as long as it is heavily involves using (deep) machine learning system. Example projects might be building a spam filter, making a neural-net based face recognition system, or building a checkers player that learns from experience with reinforcement learning and deep learning.

There are no restrictions on programming language or machine learning method.

Use of existing datasets and software packages is encouraged.

Deliverables and Deadlines

- Oct 29: Project proposal (5 points)
- Nov 5: Project readings (9 points)
- Nov 12: Detailed project plan (5 points)
- Nov 19-24: Project meeting (5 points)
- Dec 10: Final website: 10 points
- Dec 10: Final paper: 10 points

Oct 29: Project proposal (5 points)

Turn in a brief project proposal (1 page)

The proposal must be formatted with the ACM conference paper template:

https://www.acm.org/publications/taps/word-template-workflow

The only acceptable file format is Adobe Acrobat. Documents in other formats will not be accepted.

Include the following information: *(italics indicate hints and examples as to how to answer these questions)*

- Project title & group membership
- What learning problem do you want to solve? *We would like to make a weather prediction system that predicts future weather given recent observations...*
- What is the intellectual interest and practical utility of doing this project? *This is an interesting problem because….This is a useful problem to solve because….This is a challenging problem because….*
- Describe the dataset you will use to test and train your system.

  Where will you get it? Be specific

  Do you have to label it yourself or is it already labeled?

  How big is it? How many examples?

  How is it encoded?

  We plan on collecting <temperature, barometric pressure, humidity> readings taken every hour in Chicago for the last year. This is available on the

*weather.com website. Unfortunately, it is not presented as a spreadsheet of comma separated values. Instead, the user must type in an hour and a day to see those values displayed as XML for a website. We will screen-scrape this data (see a following section on how) and collect the data for every hour in the last year. This will give us 365*24 =8760 hours of data and about 8750 sequences of 10 hours that we can train and test the system on. This data will be stored in a csv file as follows*

Month	*Day*	*Hour*	*Temperature*	*Barametric*	*Humidity*
3	*9*	*10*	*45*	*30*	*.75*
3	*9*	*11*	*48*	*30*	*.70*

- Related papers to read.

  Include a minimum of 3 relevant research papers.

  Give bibliographical info.

  READ a couple of them before submitting the proposal.

  Try scholar.google.com as a starting point.

  Go back and look at the ACM template. This shows you how to format your citations. I want full bibliographic information.....AND hyperlinks to papers. A good place to find papers is scholar.google.com

Nov 5: Project readings (9 points)

This is a GROUP version of your normal readings. The twist is that you will provide analysis of 3 relevant papers that relate to your project….and that you can submit this as a group.

Nov 12: Detailed project plan (5 points)

In response to the professor's critique of your initial proposal and the readings you have done, you will submit a final detailed project plan. It will be identical in format to the initial proposal, but ideas will be more concrete, finalized and detailed. It should be **no longer than 2 pages**.

It will answer all of the questions from the initial proposal, PLUS the following additional detail.

- What is the EXACT TASK that the system will learn to do? (What is the exact input, what is the exact output?). This should be more specific than "We will predict the weather". It should be more like:

  We will predict the temperature, humidity and barometric pressure, given past weather readings. Given the statistics for the past 10 hours, the system will output a prediction for the temperature, barometric pressure and humidity on the next hour. The system will take 10 triplets, <temperature, barometric pressure, humidity>. Each triplet will represent the readings for those three values at hour 1 to hour 10. Our system will be a regression system that will predict the value at hour 11.

- What aspect of the task is your system going to optimize/improve/learn?

  This is a prediction task on a time series. We are going to use a LSTM model and the thing it will learn is to output a most-likely 11th hour's set of values, given a sequence of 10 previous readings. This will be learned CTC loss. Once the model is learned, we'll feed it a sequence of 10 readings and have it predict the most likely 11th.

Deep Learning Final Project

- What measure will you will use to evaluate performance of the system? Be specific. Don't say just "We'll measure performance."

  Our ground-truth is the actual readings (barometric pressure, temperature, humidity) at hour 11. We'll treat (barometric pressure, temperature, humidity) as 3 coordinates in a Euclidean space. We can then measure the Euclidean distance from our actual readings to the predicted values. That will be the error measure. We will average this over a large number of trials to get a mean error.

- What is the baseline approach you will compare to? Random performance? An existing system? Human performance? Be specific.

  The baseline approach we will compare to is to just assume that the weather readings at hour 11 are exactly the same as at hour 10. This is reasonable, since, lacking any other information, this is what someone would normally do. If our system can do better than this, we'll feel confident it learned something

- What software will you need to write?

  Who will write it?

  How long will that take?

  We will need to write a screen-scraper to pull the weather data down from the site weather.com. Beatrice did something similar for another class, therefore, she will be in charge of this portion of the project. The expected completion date to collect 1 year of data (24 readings per day) for our target city (Chicago) is shown in the project milestones section.

- Describe any existing software packages you will use. Answer the following questions.

  Where will you get them?

  Have you already tried them out?

  We will use Keras + Tensorflow 2. Luis downloaded it and successfully trained a model on one of their toy examples. We're confident we can use Keras..

- What is the job of each team member?
- Milestones

  What will you complete by the date of the status report?

  What will you complete by the date of the final?

  Who is responsible for each milestone?

  The table below outlines our milestones and expected completion dates.

Milestone	*Group Member(s)*	*Date*
Data collected from web	*Beatrice*	*Nov 16*
Testing software written	*Lourdes*	*Nov 20*
Baseline weather predictor done	*Larry*	*Nov 20*
Initial sanity tests of baseline & experimental setup	*Lourdes, Beatrice*	*Nov 21*
Initial Markov Model done	*Larry*	*Nov 28*
Experiments complete	*All 3*	*Dec 2*
Poster Printed	*Larry*	*Dec 7*

Deep Learning Final Project

Nov 19-24: Project meeting (5 points)

In this week your group will have a 20 minute meeting with the professor to discuss the progress of your project.

You will meet **as a group** with the professor. You are **all** expected to be **on time**. Being late or having a group member not show up, will cost points. Your grade will depend on your ability to clearly and briefly describe the high-level approach (your task, the approach, how you are measuring success, etc), as well as your ability to answer concrete questions on your progress. This is also a chance to get feedback, so use it as such.

Dec 10: Final website: 10 points

You are to build a website that gives an overview of your research project.

You are free to format the website as you see fit.

You are free to host the website wherever you like.

That said, there will be certain required elements that we will look for. The website **must** contain the following:

1. The project title (prominently displayed)
2. The name of each project member (prominently displayed)
3. At least one contact email address (prominently displayed)
4. The name of the course, university and professor (prominently displayed)
5. A 3-4 paragraph synopsis of what this work is "about"
    - Motivate the problem: (what is the thing you're trying to do and why anyone should care
    - Describe what you're doing in very high level terms
    - Describe how you built/ tested (what your dataset was, how you measured success)
    - Describe some results (how well it works in no more than a paragraph)
6. A minimum of one pretty picture or graph that illustrates your work…with a caption and to explain what the viewer is looking at.
7. A link to your final paper.

Note, these are minimal requirements. Having all these features guarantees an OK grade, not an A. Doing a good job with these things guarantees a good grade.

Note also, the plan is to make a unified final project website that contains links to all class projects. This website will be reachable from the class website and I will email the URL to the class mailing list so that you can all see each other's projects.

Dec 10: Final paper: 10 points

The final extended abstract must be **no less than two full pages and no more than three full pages**. It must be formatted with the ACM conference paper template:

https://www.acm.org/publications/taps/word-template-workflow

It should be structured like a research paper in a computer science conference.

Deep Learning Final Project

There will be certain required elements. The abstract **must** do the following:

- Motivate the problem: (What is the thing you're trying to solve? Why anyone should care? What is the intellectual interest? What is the practical utility of solving this problem?)
- Describe prior work in the area and explain why it does not address the problem
- Describe your approach in very high level terms (What kind of learner did you use? How did you apply it? What was the optimization criterion or error measure?)
- Describe how you tested and trained it (what your dataset was, how you measured success, what baseline you compared to)
- Describe some results (Are they statistically significant? What do they mean?)
- Describes where you will go from here (future work).
- A minimum of one pretty picture or graph that illustrates your work…with a caption and to explain what the viewer is looking at.

Note, these are minimal requirements. Having all these features guarantees an OK grade, not an A. Doing a good job with these things guarantees a good grade.