CSCE 823: Advanced Machine Learning

Summer 2019

Project Description

Your goal is to explore some aspect of your AFIT research using the lens of advanced machine learning / deep learning: you should frame your question as a classification / regression / clustering problem. You will produce a publication-quality document and an in-class final presentation. The deliverables are:

|  |  |  |  |
| --- | --- | --- | --- |
| **Item** | **Due Date** | **Method** | **Points** |
| Project Proposal | 10 Jul 2019 | Word/PDF via Canvas | 5 |
| Project Draft Report  (optional) | 12 Aug 2019  (first come, first served\*) | Word/PDF via Canvas | N/A |
| Project Video | 26 Aug 2019 | **Video (.mp4)** + source file (.ppt or .pptx) via Canvas | 15 |
| Project Final Report | 03 Sep 2019 | Word/PDF + code via Canvas | 25 |

The open project assignment is worth 45% of your final grade in the class.

The specifics of the assignments and the grading criteria are listed below:

## Project Proposal (5/45)

Your goal is to propose a topic to work on for your open project. Please consider the remainder of this assignment description as you develop your project proposal. Note that most projects in deep learning are assumed to be supervised learning projects – if you are planning to use something other than labeled data you should clearly indicate how you will deal with the unlabeled data in this proposal. Your project proposal should be written in standard prose (not bullet form). You may include diagrams if you wish. Size constraint: 400-1000 words (plus any graphics), single space, 10-11 point font, 0.5” margins. **PDF or MS Word required**

You should discuss:

* The domain of study you are going to work in, the specific problem / task / research question you are going to address in the project, and the reason why it is important to study it or obtain the answer to it
* A formal description of the deep learning task using language from the course - for example: spatially-organized data such as images or hyperspectral image cubes, temporally organized data such as signals or natural language; or some combination of spatial and temporal data (such as videos).
* Explain your data in detail:
  + If using existing data, do you already have it? Where did it come from? (include URL or organization)
  + If something else is being done to collect or generate the data (e.g. a simulation or an experiment), describe the experiment, and the type of data that you will be obtaining.
  + What wrangling steps will you need to do to get your data ready for ML in your project? Is it raw, semi-processed, or a complete ready-for-ML data matrix, set of labeled images, or set of labeled signals?
  + Describe the data: How many observations? How many features? Describe the features (in text).
* Where are you getting the “truth” labels from? Describe the target variable:
  + If regression, what is the range of values you are trying to predict, and what is the distribution of these values?
  + If classification, how many classes? Are the class representations in the data balanced, or unequally represented?
  + What measures of performance will you use to determine how well your ML approach works?
* How will the results of your project will support your research (or someone else’s research)
* (optional) Indicate which publication venue ( and provide venue’s URL) that you and your advisor might be interested in publishing the research in.

If you are continuing to extend the research project from CSCE 623 you must show how your work is sufficiently different. Some means of differentiating the work include applying and comparing performance of advanced techniques on the same data, and/or adding new datasets to broaden the generalizability of your previous findings in CSCE 623.

Proposal Filename/Submission details:

The name of your PDF or MS Word files should be:

Project\_proposal\_<LASTNAME\_FIRSTNAME>\_<submissionDate>

Where <submissionDate> is YYYYMMDD in numerical format

Submit via Canvas assignment link

## First Draft Paper (optional, ungraded… first come first served\*)

This is a chance to get holistic feedback on an early draft of your work. The instructor will provide feedback on any papers received in the order in which they were received. \*Note that depending on how many papers are submitted, papers submitted later than the suggested date may not receive feedback in time to incorporate suggestions prior to turning in the final paper.

By this point in your efforts you should have a great understanding of your project, including the task you are trying to accomplish, the nature of your data, what kind of neural network you will use to tackle the task, and how you will assess the quality of your fitted model and improve it. Hopefully you’ve started coding and have some preliminary results to present, and you have a good idea of what you are going to do by the end of the quarter.

In your first draft you should explain the details of how you plan to investigate your chosen topic and describe any work you’ve completed, as well as any preliminary results you’ve obtained so far.

Length/Formatting/Contents:

1000-3000 words (plus any graphics), single space, 10-12 point font, PDF or MS Word. Since this is the first draft of your paper (as opposed to the final paper), notice that some of the sections required in the final paper are not present in this description:

* INTRODUCTION: Briefly describe the problem and motivate the reader why it is interesting. State your research *questions*, *hypotheses* and/or *objectives,* which should have some *theory* behind why you believe them to be true (usually theory comes from the literature review). Describe the details of your problem space. You should try to address as many of the problem definition details (see below) as possible
  + Example Research Question: “Does data augmentation using the <insert type here> technique help improve performance on the <dataset name> dataset?”
  + Example Research Hypothesis: “Algorithm A achieves a 30% improvement over B according to the … measurement … when the we add the <technique name> ”
  + Example Research Objective: “To develop a <> model which uses features X1 … to determine operator workload in real time with accuracy = …, and F-measure = ….”
* RELATED WORK: Identify and briefly discuss any references you’ve found and used/plan to use so far. You should be using your course text at a minimum, and you **must include at least 2 other published papers** (conference or journal), which do not include references which just describe your dataset. Consider including reference papers you have read for your research if relevant. Look for *theory* to include here. For example, it is believed that the X1 brainwaves increase while the X2 brainwaves decrease when an operator is experiencing fatigue. Try to include work which “competes” with yours proposed project… what papers cover the state-of-the-art performance in this area?
* DATASET: Describe the type of data you are using, where it was obtained from (or how you generated it) with any preprocessing steps required before it can be used in machine learning algorithms. Include sources of the data (website/papers). If your dataset is small, what are your plans to obtain/create enough data?
* APPROACH/METHODOLOGY: How are you solving the problem? Which types of neural networks (and layers) will you use, and why? Try to relate your approach to something discussed in class or your reading. Include details of the experiments you have set up or plan to run, and the analysis you plan to perform. Describe the form of the results that you plan to obtain from such experiments/analysis. (for example: if your experiment uses classification for a 2-class problem, you might want to make a ROC curve; multiclass will be served well by a confusion matrix). Describe details of your experiment settings, including parameters, code versions, hardware configuration & OS, etc. The goal is to ensure a future researcher could duplicate your work and convince your reader of the level of rigor you used. Show how your results will have statistical rigor and describe the ways in which you minimized estimation error in your results (are you using confidence intervals, T-tests or Z-tests, cross-validation, etc.)

When detailing your Approach / Methodology and Results also make sure you address these:

* What are the measures you are using to determine the quality of your solution?
  + How will you evaluate performance results – in aggregate? An investigation of residuals?
  + Are you comparing performance to some baseline, or compare to the naïve/null alternative?
  + Are you trying to rank-order the quality of several alternatives?
* If your system uses separate training & testing phases (i.e. cross-validation), what is your technique, and what parameters are you using (for example: the value of k in k-fold cross validation)?
* (EXPECTED) RESULTS & ERROR ANALYSIS: If you have any preliminary results, put them here. Show how you looked at the errors your network is making to see if they inform a better architecture or training methodology. Also, describe how the line of research, experiments or analysis you devised, and representation of the results you chose would serve to answer the research question you asked or how they reject or fail to reject the null hypothesis you set up
* REFERENCES: Include a bibliography section with a **minimum of 3 references** beyond the reference(s) for your dataset – examples (book + 2 papers).

First Draft Filename/Submission details:

The name of your PDF or MS Word files should be:

Project\_firstDraft\_<LASTNAME\_FIRSTNAME>\_<submissionDate>

Where <submissionDate> is YYYYMMDD in numerical format

Submit via Canvas assignment link

## Project Videos (15/45)

**Project Video:** Your goal is to present your research to your classmates in a succinct format. The file format must be **.mp4** Your video presentation should be approximately 4-6 minutes total. Too-long or too-short videos will not receive full credit. Your video will be graded by your instructor on your presentation completeness & correctness, and pace/quality – to include total length of the video.

Your video should strive to cover the key points of *your* work and should have good pace and high quality:

|  |
| --- |
| MOTIVATION/RESEARCH QUESTION: What is the problem you are working on & why is it important? What is your research question / hypothesis / or objective? Include some background on your specific problem domain. |
| DATA EXPLANATION: What is the nature of the data (or experiments that generate the data) you are working with? |
| APPROACH/METHODOLOGY: What was the approach you used in your research? What did you apply from class or your reading to address the issue? |
| ANALYSIS TECHNIQUES: How did you set up and ensure rigor in your research - provide rationale/justification that your technique was valid |
| RESULTS: What were your results? Show some graphs/tables/figures and describe them. |
| CONCLUSION: What are the impact of your results, and what assumptions/limitations are there with your results |
| FUTURE WORK: What would you work on next if you had more time (how would you go about removing the assumptions/limitations of your work) and how will this be used in your AFIT research |
| PACE (4-6 minutes; adequate speed for audience) |
| QUALITY (audio & video) |

You must submit a .mp4 file (and a source file if you are using powerpoint). While you are free to use whatever software you want to create your video, one easy way is to use powerpoint narration mode which produces a PPT file. If you use powerpoint to create your video, you should also submit the powerpoint file (pptx). Preview your video before submission: Watch your video on a computer other than the one you created it on, without using headphones. Ensure the video and audio quality are adequate. You may also want to get another person to watch your video and give you suggestions before you submit it. Things to watch out for when you are developing your video:

* Sound quality: Beware of both too low volume and clipping/overdriving the audio by talking too loud or too close to the microphone. Avoid recording in noisy environments (or turn off fans or A/C). Minimize sounds of mouse clicks, keyboard presses, tapping, or rustling papers. Fix very soft or very loud segments. The sound level should be easily heard at midrange volume (its easier to turn down the volume than to crank up a weak signal after you finalize the video). If you are recording narration in multiple segments be consistent with the location, speaking volume and distance from the microphone (or adjust the volume before finalizing the video)
* Image/picture quality: avoid blurry image captures. Avoid small fonts on the axis/legends on graphs – your audience needs to be able to read everything on the screen, even after video compression.
* Ensure your timing & slide changes match your narration end times in the final video – watch out for narration cutoffs which seemed fine before you encoded the video or long wait times after the end of the audio before the next slide.

Presentation Filename/Submission details:

Include the final video and the source file used to create it. The name of your files should be:

Project\_video\_<LASTNAME\_FIRSTNAME>\_<submissionDate>.mp4

Project\_source\_<LASTNAME\_FIRSTNAME>\_<submissionDate>.PPT or .PPTX (if you created with powerpoint)

Where <submissionDate> is YYYYMMDD in numerical format

Submit via Canvas assignment link

## Project (Final) Paper (25/45)

Your goal is to create a complete publication-quality paper (conference level or better) describing your research. You must make sure that all of the components of the Draft Paper report are covered and expanded to include the progress you’ve made since the draft. You will be graded on the completeness of your document (15/25) as well as the quality of your writing (5/25). Use the paper checklist to guide your effort and ensure you maximize the completeness and correctness of your work.

If you have a specific target journal or conference that you are planning to submit your paper to, you should use the style and format guidance provided by the venue (please indicate which venue, provide the hyperlink, and indicate the submission deadline in your assignment submission notes in Canvas). Otherwise, use the following guidance:

Length/Formatting/Contents:

2000-4000 words (not including references/bibliography), double space, 10 point font. PDF or MS Word accepted. Your final draft should also add or update the following sections:

* ABSTRACT: In 150 words or less, cover the gist of your paper - succinctly describe and motivate the problem you are working on, touch on your approach and list your key findings or results, along with why they are an important contribution for the field.
* INTRODUCTION: as described in the draft paper description
* RELATED WORK: as described in the draft paper description
* APPROACH/METHODOLOGY: Update this section by describing any additional details of your process description for statistical analysis and include any settings you used in your statistical measurements (i.e. alpha in your hypothesis test).
* RESULTS: Update this section by discussing your final results. Include and explain any graphs/tables in the main text. Make sure to use labels and captions to properly reference & briefly describe figures/tables/graphs in your work.
* CONCLUSION & FUTURE WORK: Re-iterate how the line of research, experiments or analysis you devised, and representation of the results you chose serves to answer the research question you asked and how it confirmed or refuted your hypothesis, met your objective, or answered your question. Describe any assumptions/limitations of your research that you would like to see removed in future work, and suggest at how one might go about removing them.
* REFERENCES: as described in the draft paper description

Final Written Report Filename/Submission details:

The name of your PDF or MS Word files should be:

Project\_finalDraft\_<LASTNAME\_FIRSTNAME>\_<submissionDate>

Where <submissionDate> is YYYYMMDD in numerical format

Submit via Canvas assignment link