COMP 347	Applied Machine Learning

Final Exam (Project)

Assigned: November 14   *Due: December 10 3:00 PM*

Abstract

In this project you get to learn about a machine learning topic we haven't covered and teach it to the class. It will probably surprise you how much more deeply you learn something when you have to teach it. You have a lot of flexibility in the topics you can choose.

Task

For this research project you will need to pick a machine learning topic, learn about it from multiple sources, and give a ~20 minute presentation to the class. Your goal is to provide a self-contained overview of the topic that also teaches something meaningful – do not just read a bunch of definitions off of some slides!

Presentation Requirements:

- Approximately 20 minutes, with the expectation of about 5 minutes of questions. Knowing the answers to audience questions is hard – you will need to study the material beyond what you present.

- Some kind of instructor aide that will be turned in to me. This might be powerpoint slides, or simple handwritten notes that help guide your "lecture".

- At least one example and/or exercise, but two (or more) is always better.

In addition to the presentation, you will need to turn in (1) an annotated bibliography, (2) a Jupyter notebook with some code that applies your chosen topic, and (3) a review of each of the other presentations.

Picking a Topic

There are so many potential topics! Some obvious choices are any models or model variants that we didn't get to talk about, e.g., newer neural network architectures, the OPTICS algorithm for clustering, etc. There are many advanced variants of the models we did talk about. There are lots of options in scikit-learn, and even more in other packages – explore what's available!

You could go deeper into something that we touched on, such as the math for a particular model, good ways to visualize and/or interpret models, or feature selection.

You could also go into a subtopic that we didn't cover at all, e.g., semi-supervised learning, active learning, reinforcement learning, and more.

You are not limited to these examples. Do some searching and be creative. If your topic is sufficiently broad you might need to pick a particular focus within it.

Officially Selecting a Topic

You must inform me of your topic via email (see the "Deadlines" section). Topics are first-come, first served, so you might not get the topic you want if someone else picks it first.

Finding Sources

If your topic is included in the textbook, you can start there. The textbook provides many references to other sources.

Aside from that, web sources are fine, although published papers are better than random blog posts. Wikipedia can provide a great jumping-off point, since those pages are generally loaded with references.

Annotated Bibliography

When you have gathered enough sources (at least 4), you need to submit an annotated bibliography. This would include all of the sources you believe you'll use for your presentation/study (though it is ok to add more later). Annotations should be no more than a paragraph, should give a high-level overview of the source, and highlight its importance and/or relevance to your topic and focus area.

Making a Presentation

In addition to learning the material, don't forget to spend some time making a good presentation. Slides can help keep you on task, but do NOT just read from your slides!

There are lots of resources online about making good slides, or presentations/lectures in general. If you're making slides, don't overload slides with text. Try to aim for 1 figure for every slide. You may not always hit this goal, but its good to shoot for. If you can't think of a figure for a slide, it may mean it needs to be combined with another slide, split into multiple slides, or that you don't understand that concept well enough.

In any case, you will definitely want an outline. Present this outline near the beginning of the presentation (either right away or perhaps after a motivating example). Don't be afraid to remind the audience of your outline at various points.

Don't forget about motivation. After you get deep into a topic, you start to take for granted why people should care about it, and it is easy to launch into a lot of details. Details are fine and necessary, but first you need to make sure the audience understands WHY this concept exists – what problem is being solved, what are the other solutions, and why is this particular one good and bad. It's easy to forget that last one – if you can't explain the downsides of a particular protocol/topic, you probably don't understand it. Everything has tradeoffs.

Lastly: examples, examples, examples. It is hard to come up with examples and still cover the material, and it is harder still to work through an exercise with audience input. But it is the best way to learn.

Jupyter Notebook

In the process of teaching/explaining your topic, I expect that you will show some code snippets applying your topic.

Along with your presentation, you need to submit a Jupyter notebook that contains some code that applies your topic to a dataset.

Peer Reviews

You are expected to be attentive and respectful during the presentations. I highly encourage you to ask questions. Having questions at the end of a presentation is a sign of a good presentation; they

should not be seen as a criticism. You can ask about things that are unclear, or just any real-world implications for the topic, e.g., can this be used for X?

For each presentation (other than your own), you will need to provide a peer review. Peer review forms will be provided for you.

Timeline

Presentations will occur on our final exam day – Saturday, December 10.

Date	Checkpoint
11/16	Topic claiming period opens
11/22	Topic must be approved
12/02	Annotated Bibliography due
12/10	Presentations