CS 449 Final Project Proposal

Due: November 26, 2024 at 11:59pm

If you are using this template to write your update, please delete the instructions and examples before you turn in the PDF you want graded, including this very sentence.

1 Names

List all members in your group.

2 Abstract

Write a new abstract that describes the motivation for your project and proposed methods. This should reflect any changes you've made, but you'll explain those changes later.

For example:

Our final project seeks to use a variety of `sklearn` models to classify handwritten digits in the `MNIST` dataset. We will compare models such as Logistic Regression and Multilayer Perceptrons.

3 Big Changes

Describe any major changes that have taken your project in a new direction. What made you decide to make these changes? Were some things easier or harder than you expected?

For example:

We intended to train sklearn models on the MNIST dataset, but found that was way too easy. We instead decided to use torch models on the more complicated Fashion MNIST dataset. This led us to an interesting challenge where...

4 Data

4.1 Describe your dataset(s)

This should be an updated version of your datasets section from your proposal. Write this section as if the reader hasn't read your original proposal. List the datasets you're using, where you found them, and what they contain. Be detailed! For each dataset, what does the data look like? What is the data representation? (e.g., what resolution of images? what length of sequences?) How is the data annotated or labeled? Include citations for the datasets. Include at least one citation of previous work that has used this data, or an explanation of why it hasn't been used before.

For example:

We use the popular Fashion MNIST dataset, which contains ... greyscale images of clothing. The dataset was created in ...

4.2 Load your dataset(s)

Demonstrate that you have made at least some progress with getting your dataset ready to use. Load at least a few examples and visualize them as best you can

4.3 Small dataset

Describe the small dataset you constructed as requested in the project proposal. How many examples did you use? How large was the dataset in total?

For example:

We used the first 1000 examples from our dataset and downsampled our images to be 8x8 pixels. . .

5 Methods

This should be an updated version of your methods section from your proposal. Write this section as if the reader hasn't read your original proposal. This section should be more detailed than it was in your proposal, since you've made more progress by now.

For example:

This is a standard supervised learning task, and we will use 'sklearn' and 'torch' models to predict what kind of clothing appears in small images. The 'sklearn' model will contain one weight per pixel, but our 'torch' models will contain millions of weights. We will train our model using Cross-Entropy loss, because...

6 Old Deliverables

Copy the deliverables from your revised proposal and write a sentence saying whether you completed this goal, made progress on this goal, or abandoned this goal. Unless you completed the goal as you expected, give an explanation of how it went – what was harder than expected?

6.1 Previous Essential Goals

Include all your essential goals here

For example:

We will use Logistic Regression, Support Vector Machine, and Decision Tree models to train and test on our MNIST data.

Update: We completed this goal by using the 'sklearn' packages.

6.2 Previous Desired Goals

Include all your desired goals here

For example:

We will conduct a thorough hyperparameter search for all our models and provide a detailed and practical guide for how a researcher should search for good hyperparameter values.

Update: We made progress on this goal by running a hyperparameter grid search for both sklearn and torch models, but...

6.3 Previous Stretch Goals

Include all your stretch goals here

For example:

We will conduct a manual analysis of the digits that our model gets wrong and use a GAN to create new images that help us learn a more robust classifier.

We abandoned this goal because we didn't want to conduct a manual analysis.

7 Results So Far

Show the results of your methods trained on your data. You should at least have empirical results for the small dataset described in 4c. Unless you are doing generation or another task where simple models (e.g., linear regression or logistic regression) aren't applicable, we strongly encourage you to compare your initial results against a simple baseline. If you have additional results, you're welcome to include those as well.

8 New Deliverables

For any proposed deliverables that you did NOT complete or abandon, copy them into these lists below. Then, if you have abandoned or completed your Desired or Stretch goals, add new goals until you have at least two goals per category.

Make sure you have two remaining goals in each of the Desired and Stretch categories.

9 Hopes and Concerns

What's gone well so far? What's been the hardest thing?

10 References

Cite the papers or sources that you used to discover your datasets and/or models, if you didn't include the citation above.