Independent Study Research Idea Proposal

Kyle Remmenga

The goal of this independent study is to understand the architecture behind a large language model by studying and developing a small language model. Large language models require immense resources both in terms of hardware and datasets thus by implementing concepts that exist within small language models, a strong understanding of the architecture, mechanisms and optimizations that are necessary to develop a modern language model could be achieved. In this way, we are able to mock a large language model in terms of its structure and algorithms in order to create a learning environment that could be applied to larger projects as well as other machine learning applications in the field of computer science. With this idea in mind, this proposal seeks to outline a curriculum for an independent study that would allow a student to delve into the mathematics and algorithms that are currently seen as the cutting edge in the study of machine learning.

To start, the curriculum would focus on developing an understanding of the architecture behind neural networks. This would include creating a foundational understanding of feedforward networks and backpropagation. From there the study would shift to activation functions and loss functions and their applications within language models. Lastly the curriculum would shift to cover optimization algorithms with a focus on gradient descent algorithms. It is the hope that after these topics are covered the student is able to develop a simple feedforward neural network using NumPy. The student could then incorporate a basic data set to train the model to classify images.

The next part of the curriculum would move to build an understanding of language modeling and N-gram models as well as recurrent neural networks. After this, the student would be tasked to understand long short-term memory and gated recurrent units and lastly this section would conclude on the limitations of recurrent neural networks such as vanishing/exploding gradients. After this section the student could be tasked with developing an implementation of a recurrent neural network and train this model on a small text dataset to predict the next character in a sequence.

This section of the curriculum will focus on attention and its value within machine learning projects. This section will start with building an understanding for a neural network's need for attention. Following this the focus will shift to an introduction to attention mechanisms and build on this to cover additive and multiplicative attention. Lastly this section will cover the self-attention concept and its application in optimizing neural networks. After this section the student could attempt to enhance the previous recurrent neural network model by incorporating attention. The student should be able to visualize attention weights to interpret the model's focus.

The next section of the curriculum will focus on transformer architecture and will start on a deep dive into the "Attention is All You Need" research paper. From here this section will explore transformer components such as multi-head self-attention, positional encoding, layer normalization and residual connections. To conclude this section the student will research encoder and decoder structures. After this section the student should be able to implement a basic transformer encoder layer from scratch then compare its performance with earlier recurrent neural network based models.

At this point in the semester all sections will have been covered and the student will then focus entirely on their paper and implementations. The paper seeks to analyze how different techniques impact the outcome of the model as well as the effectiveness of various optimization

strategies. The student's paper will be of publication standard and will detail their implementations throughout the semester and how different architectures affected the model. In addition the paper will include the process of optimization and how effective these different techniques were in improving the model. The student will also attempt to implement a small language model. The purpose for this implementation is to help the student understand the architecture and mechanisms within the model therefore the expectations for the quality of this model will be undefined.

This curriculum seeks to describe key elements of a large language model as well as other machine learning models to allow a student to research these elements in isolation as well as in combination to build an understanding for how these models work. The curriculum attempts to break down elements so that the student can learn how they work and analyze their effectiveness when used together in machine learning architecture. In addition, this would allow the student to build upon their understanding by comparing different strategies and analyze how these components work together in the larger framework. Furthermore the student would learn how to compare machine learning models and why that is necessary for the advancement of this field of study.