HW5+6: Neural network experiments + current state of the art of AI

This homework requires you to: * perform experiments with a neural network to search for the best architecture for a classification task * read and comment on three key papers delineating the current state of AI

Finding the optimal architecture for a neural network (5pts)

- Visit https://playground.tensorflow.org
- Turn on "Show test data" check box.
- Set the batch size to 10, ratio of training to test data to 50%.

Problem 1

Experiment with various datasets and architectures to get the feel of how training is working, and how the activation function, learning rate, number of hidden layers and nodes in the hidden layers affect the outcome.

Write up your conclusions in two paragraphs.

Problem 2

Set the data to be be Spiral, Noise=25. We consider the learning "successful" if the test loss is less than 0.1. Through experimentation find the neural network architecture that successfully learns to classify this dataset.

Write up your conclusions in two paragraphs.

(2017) Attention is all you need (5pt)

Read the paper: "Attention is all you need" by Vaswani et al. This paper introduced the Transformer component, the fundamental building block of the neural network architectures that are considered state-of-the-art in the early 2020s (for instance, all large language models). The paper was cited about 100,000 times.

Questions:

Write a 400 word essay about the paper. In it:

- Summarize the overall claims in the paper. Why is the title surprising?
- Summarize the explanations made in section 4.
- Explain and interpret the additional figures on pages 13-15.

(2019) BERT (5pt)

Read the paper "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" by Devlin et al. This paper triggered the overall transition of natural language processing research towards the large language model paradigm, as many NLP tasks that previously were considered to need specific architectures were shown to be solvable very easily with BERT. The paper was cited about 84,000 times.

Questions

Write a 400 word essay about the paper. In it:

- Discuss the difference between pre-training and fine-tuning BERT (Figure 1).
- Discuss the way the input is represented (Figure 2)
- Discuss how BERT can be adapted to different tasks (Figure 4 in Appendix)

(2020) GPT-3 (5pt)

Read the paper "Language models are few-shot learners" by Brown et al. This paper introduced GPT-3, raising the performance expectations of large language models. In contrast to previous papers, it also discusses some of the considerations of fairness, misuse etc, which are a significant part of the current conversation on AI. The paper has about 17,000 citations.

Questions

Write a 400 word essay about the paper. In it:

- Discuss what performance advances are claimed, and what innovations led to these advances.
- Why do you think that this paper has less citations than the Transformer and BERT papers?
- Briefly summarize the 7.1 Misuse of Language Models section.