CS 7150: Deep Learning — Spring 2021 — Paul Hand

Day 12 — Preparation Questions For Class
Due: Wednesday 3/3/2021 at 2:30pm via Gradescope

Names: [Put The Names Of Your Group Here]

You may consult any and all resources in answering the questions. Your goal is to have answers that are ready to be shared with the class (or on a hypothetical job interview) as written. Your answers should be as concise as possible. When asked to explain a figure, your response should have the following structure: provide context (state what experiment was being run / state what problem is being solved), state what has been plotted, remark on what we observe from the plots, and interpret the results.

Submit one document for your group and tag all group members. We recommend you use Overleaf for joint editing of this TeX document.

Directions: Read 'Attention Is All You Need' (Transformers).

- Read the entire paper

Question 1. *Explain scaled dot-product attention.*

Response:

Question 2. *Explain multi-head attention.*

Response:

Question 3. *Explain the positional encoding that was used.*

Response:

Question 4. *Explain the transformer architecture.*

Response:

Question 5. *Compare and contrast transformers with recurrent neural networks.*

Response:

Directions: Read 'Language Models are Few-Shot Learners' (GPT-3).

- Read Section 1, 2, 3.0, 3.1

Question 6. *Explain Figure 1.2*

Response:
    Context:
    What is plotted:
    What we observe:
    Interpretation:

Question 7. *Explain Figure 3.1*

Response:
    Context:
    What is plotted:
    What we observe:
    Interpretation: