

CS 7150: Deep Learning — Fall 2024 — Paul Hand

Day 8 — Preparation Questions For Class

Due: Tuesday 10/8/2021 at 12:30pm 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 ‘[Deep Learning Book - Chapter 8](#)’ by Goodfellow et al.

- Read Section 8.3.2, 8.5.1, 8.5.2, 8.5.3, 8.5.4.

Question 1. *What is the idea behind momentum in first order optimization?*

Response:

Question 2. *What is the idea behind the AdaGrad algorithm?*

Response:

Question 3. *What is the idea behind the RMSProp algorithm?*

Response:

Question 4. *What is the idea behind the Adam algorithm?*

Response:

Directions: Read ‘Adam: A Method for Stochastic Optimization’ by Kingma and Ba.

- Read Section 1, 2, 5, 6.3

Question 5. *The authors say that the benefits of Adam are: that the magnitudes of parameter updates are invariant to rescaling of the gradient, its stepsizes are approximately bounded by the stepsize hyperparameter, it does not require a stationary objective, it works with sparse gradients, and it naturally performs a form of step size annealing. Briefly interpret/explain each of these points.*

Response:

Stepsizes are approximately bounded by stepsize hyperparameter:

Does not require stationary objective:

It works with sparse gradients:

Naturally performs a form of step size annealing:

Question 6. *Explain Figure 3.*

Response:

Context:

What is plotted:

What we observe:

Interpretation: