MSc. Data Science & Artificial Intelligence

Advanced Deep Learning

Dr Alessandro BETTI

Assignment 1

Author: Joris LIMONIER

joris.limonier@gmail.com

Due: October 5, 2022

Contents

List of Figures

1 Exercise 1

2 Exercise 2

2.1 Question 1

I was not familiar with the PyTorch library, so I had to perform some research in order to know how its methods/classes work. I consulted several sources, including the following, which were useful in order to answer this question:

1. https://discuss.pytorch.org/t/how-sgd-works-in-pytorch/8060/2

2. https://discuss.pytorch.org/t/performing-mini-batch-gradient
   -descent-or-stochastic-gradient-descent-on-a-mini-batch/21235

3. https://pytorch.org/docs/stable/generated/torch.optim.SGD.html

According to those sources, PyTorch's SGD actually computes a full-batch (vanilla) Gradient Descent, based on the data that is passed to it. It is my understanding that in order to perform actual mini-batch (*i.e.* where $1 <$ batch size $<$ number of observations), one simply needs to give subsets of the data a each iteration.

In our case, we use the full dataset in `outputs = net(inputs)`, which is why we perform full-batch GD, although we call the `optim.SGD` class.

2.2 Question 2

Figures 2 and 1 show a comparison between train and test cross entropy loss. The leftmost diagrams is with a linear scale for the vertical axis, while the rightmost diagrams have a logarithmic scale for the vertical axis. Beware that the number of epochs varies significantly between figures. Indeed, more pararmeters in the network implies longer training times per epoch, which in turn means that running for the same time both experiments results in different number of epochs.
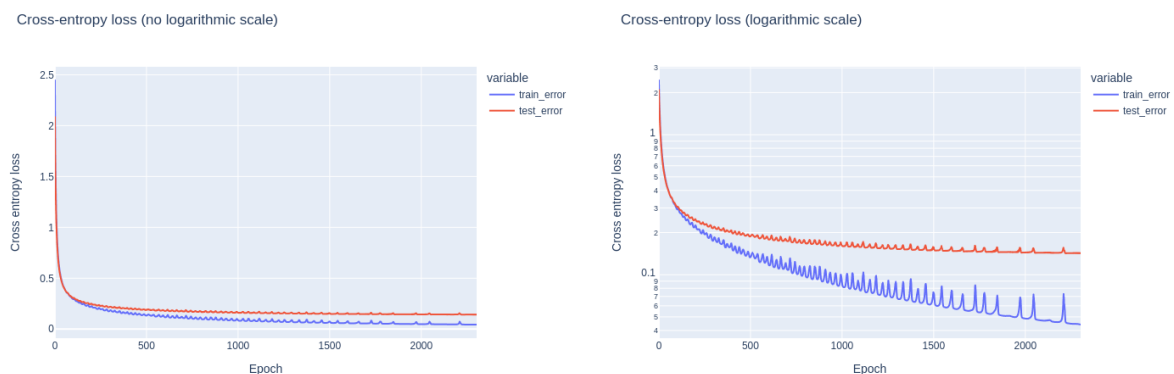

Figure 1: Cross entropy loss, train vs test (300 fully-connected).

We see that in both experiments the train error keeps decreasing, while the test error stays constant. This means that our network is overfitting and there is no need to continue
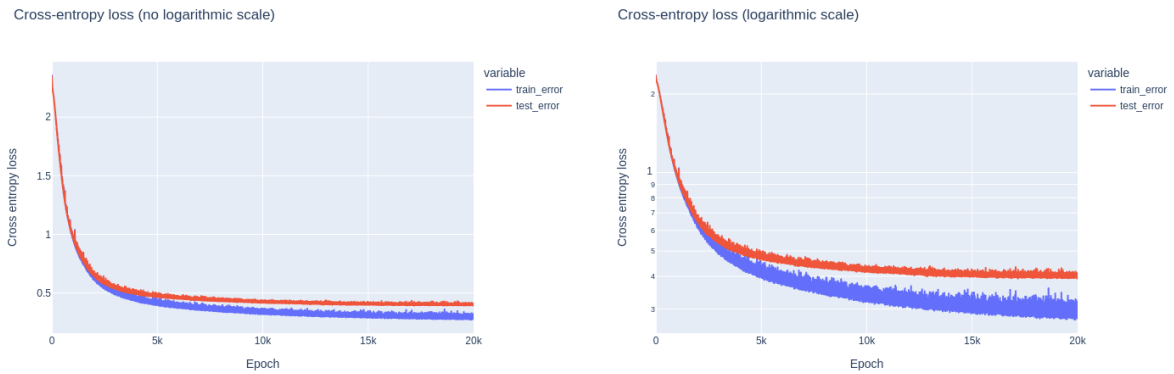
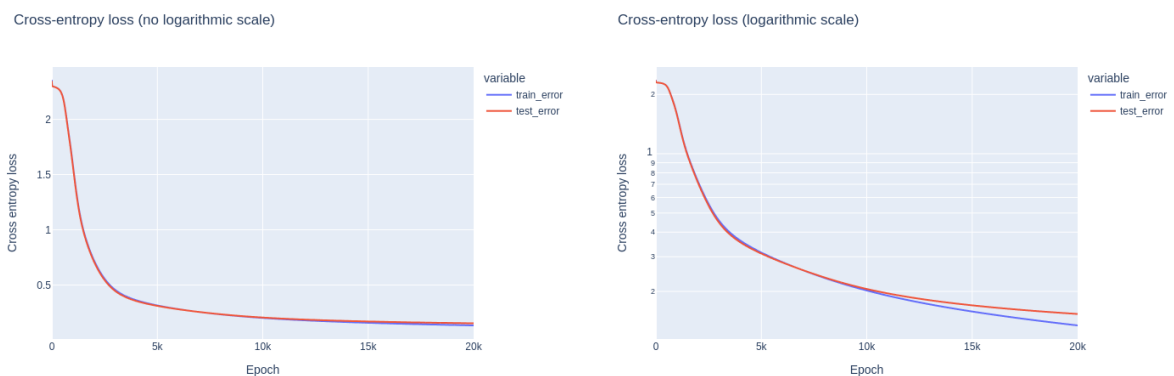Figure 2: Cross entropy loss, train vs test ($10 - 10$ fully-connected).

Figure 3: Cross entropy loss, train vs test ($20 - 20$ fully-connected).

the experiment further. One way to prevent wasting time training when the network isn't making actual progress is to print the loss, or even to plot the loss curves, every couple iterations.

2.3 Question 3

2.4 Question 4

2.5 Question 5

2.6 Question 6

2.7 Question 7