

Homework 1: Autoregressive Models

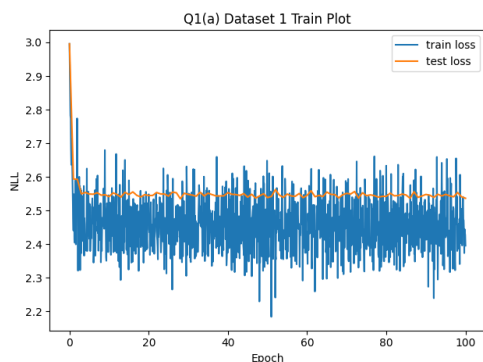
Name: Eli Bronstein
Student ID: 26086997

Deliverable: This PDF write-up by **Tuesday February 7th, 23:59pm**. Your PDF should be generated by simply replacing the placeholder images of this LaTeX document with the appropriate solution images that will be generated automatically when solving each question. The solution images are automatically generated and saved using the accompanying IPython notebook. Your PDF is to be submitted into Gradescope. This PDF already contains a few solution images. These images will allow you to check your own solution to ensure correctness. Submit this PDF, your iPython notebook, and any other code you wrote to Gradescope!

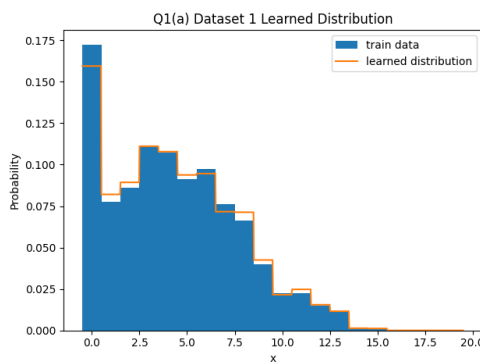
Question 1: 1D Data

(a) [10pt] Fitting a Histogram

Final test loss for dataset 1: 2.5365 nats / dim

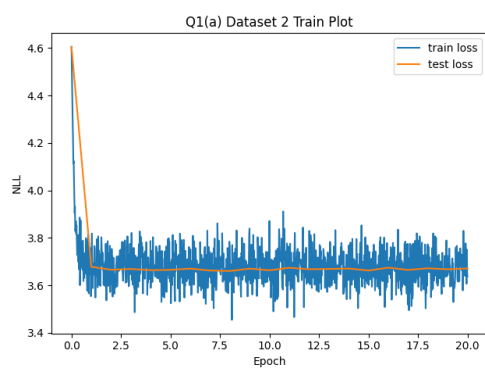


(a) Dataset 1: Training curve

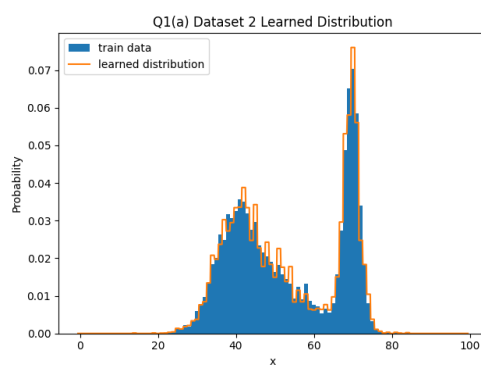


(b) Dataset 1: Learned distribution

Final test loss for dataset 2: 3.6712 nats / dim



(a) Dataset 2: Training curve



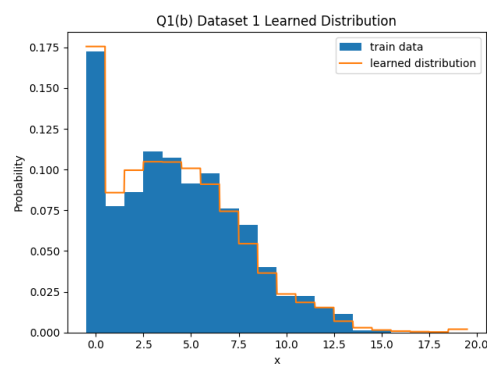
(b) Dataset 2: Learned distribution

(b) [10pt] Fitting Discretized Mixture of Logistics

Final test loss for dataset 1: 2.5495 nats / dim

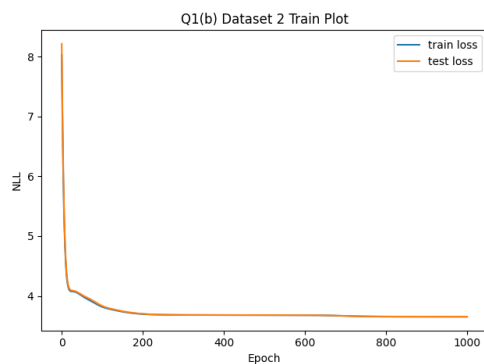


(a) Dataset 1: Training curve

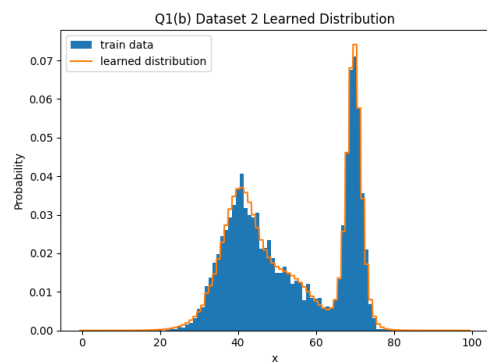


(b) Dataset 1: Learned distribution

Final test loss for dataset 2: 3.6542 nats / dim



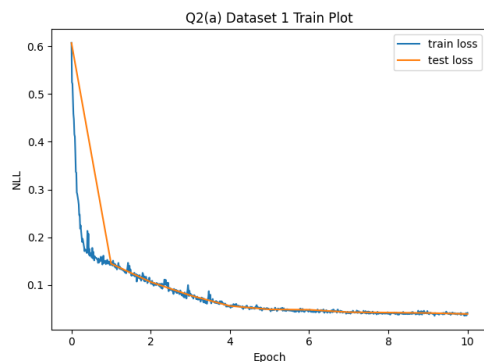
(a) Dataset 2: Training curve



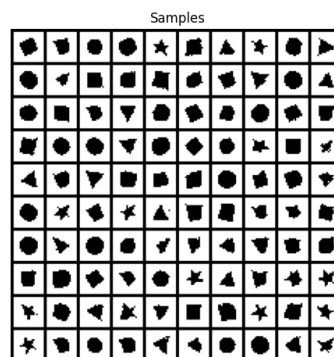
(b) Dataset 2: Learned distribution

Question 2: PixelCNNs**(a) [15pt] PixelCNNs on Shapes and MNIST**

Final test loss for dataset 1: 0.0396 nats / dim

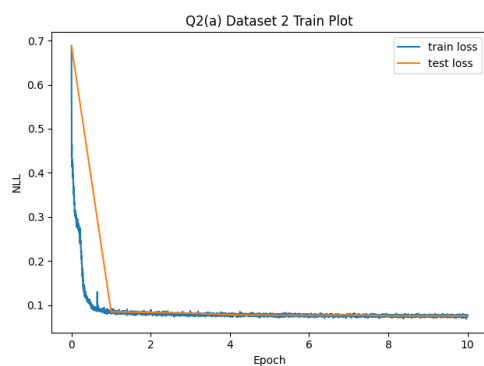


(a) Dataset 1: Training curve

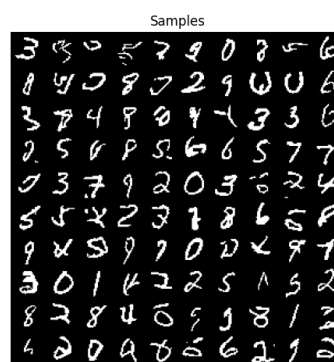


(b) Dataset 1: Samples

Final test loss for dataset 2: 0.0757 nats / dim



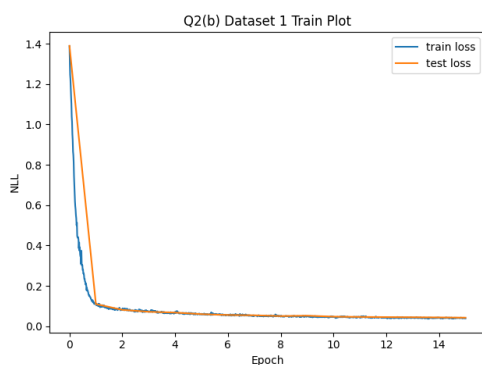
(a) Dataset 2: Training curve



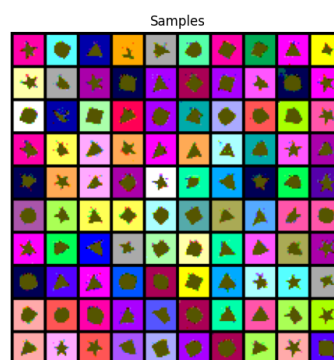
(b) Dataset 2: Samples

(b) [15pt] PixelCNN on Colored Shapes and MNIST: Independent Color Channels

Final test loss for dataset 1: 0.0425 nats / dim

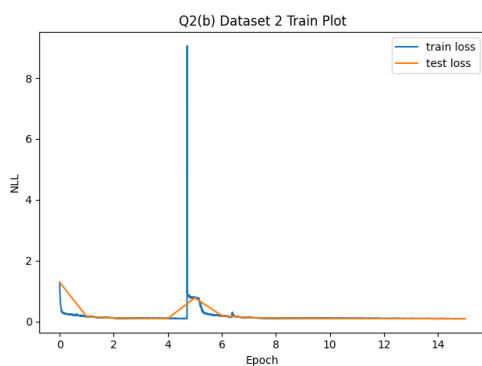


(a) Dataset 1: Training curve

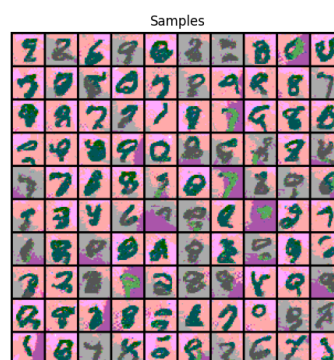


(b) Dataset 1: Samples

Final test loss for dataset 2: 0.0892 nats / dim



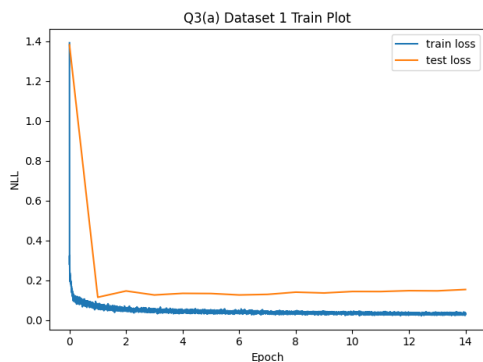
(a) Dataset 2: Training curve



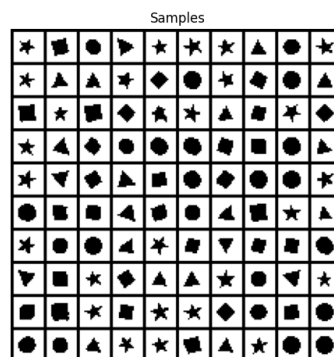
(b) Dataset 2: Samples

Question 3: Causal Transformer - iGPT**(a) [15pt] Autoregressive Transformer on Shapes and MNIST**

Final test loss for dataset 1: **TODO: update this. Should be around 0.0397.** nats / dim

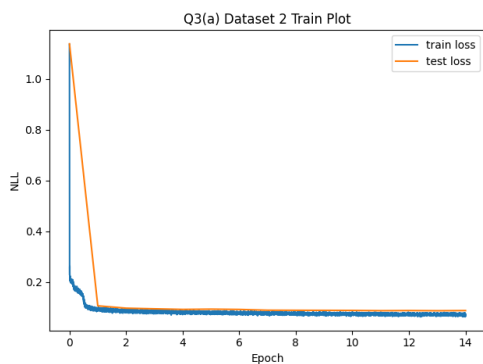


(a) Dataset 1: Training curve

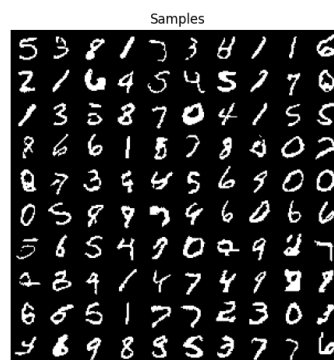


(b) Dataset 1: Samples

Final test loss for dataset 2: 0.0881 nats / dim



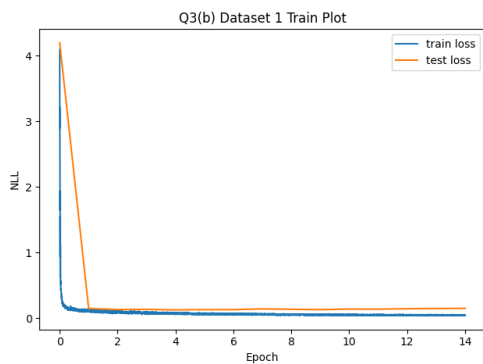
(a) Dataset 2: Training curve



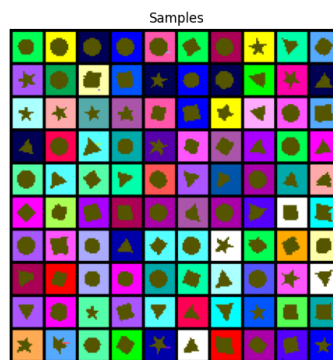
(b) Dataset 2: Samples

(b) [15pt] Autoregressive Transformer on Colored Shapes and MNIST

Final test loss for dataset 1: **TODO: update this.** Should be around 0.0541 nats / dim

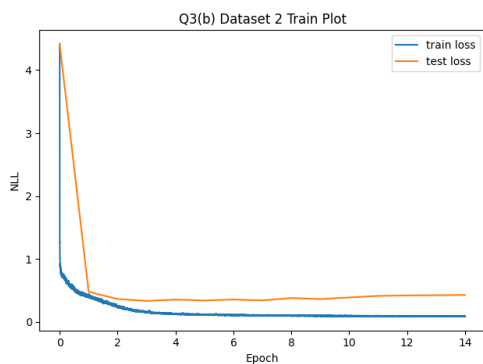


(a) Dataset 1: Training curve



(b) Dataset 1: Samples

Final test loss for dataset 2: **FILL IN HERE** nats / dim



(a) Dataset 2: Training curve



(b) Dataset 2: Samples

(c) [15pt] K,V Caching for Improved Inference

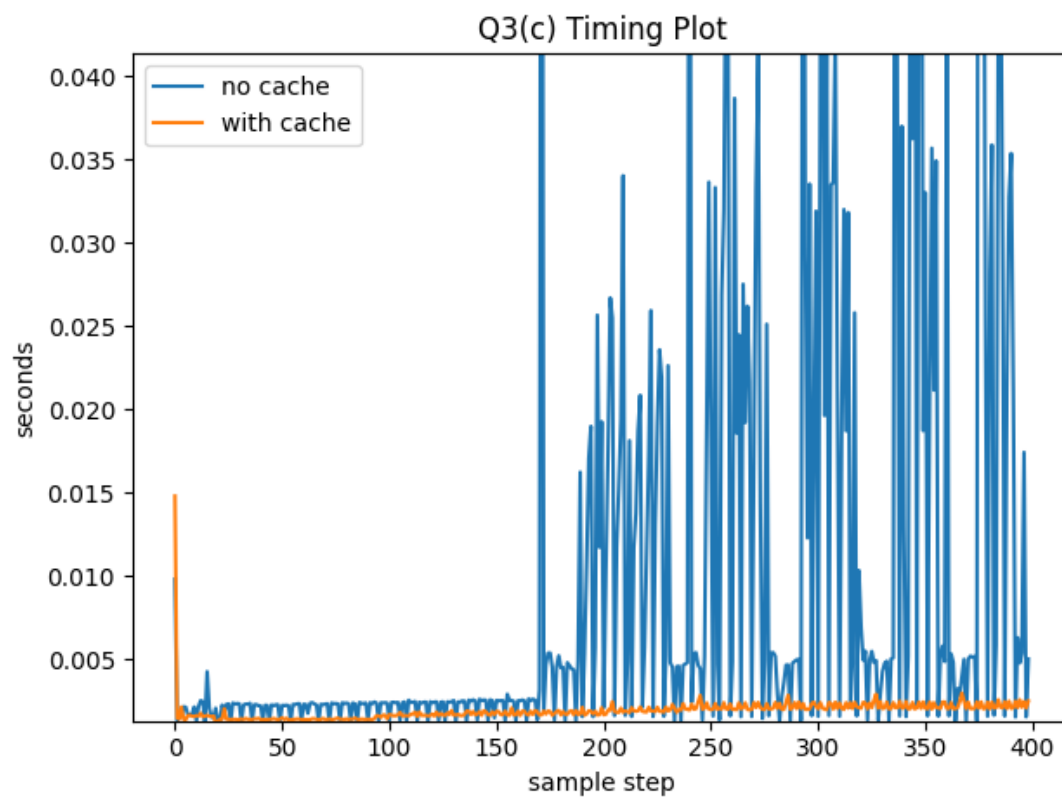
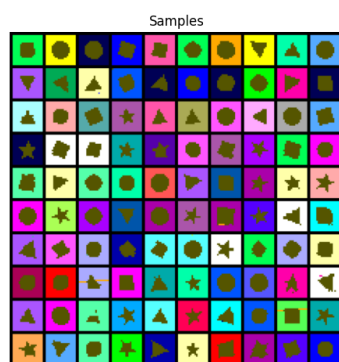
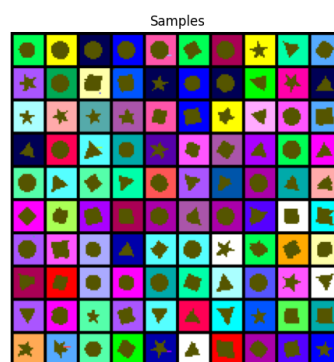


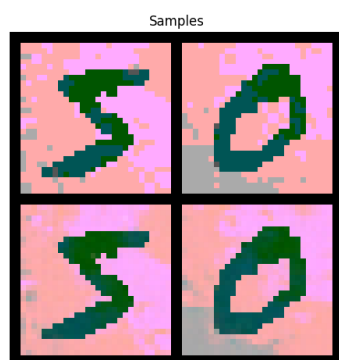
Figure 13: Dataset 2: Inference Speed



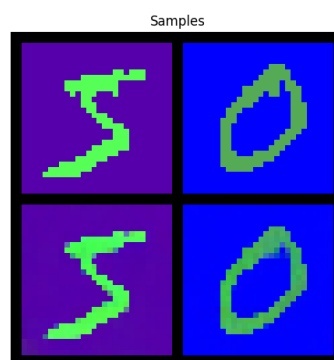
(a) Dataset 2: Samples (no caching)



(b) Dataset 2: Samples (caching)

Question 4: Causal Transformer - Tokenized Images**(a) [5pt] Image Quantization**

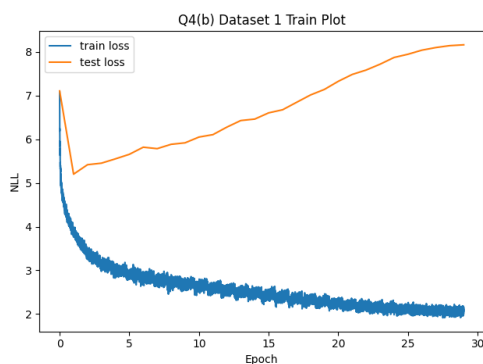
(a) Dataset 1: Quantized Examples



(b) Dataset 2: Quantized Examples

(b) [15pt] Autoregressive Transformer on Colored Shapes and MNIST with Vector Quantization

Final test loss for dataset 1: **Update this. Should be around 3.083 nats / dim**

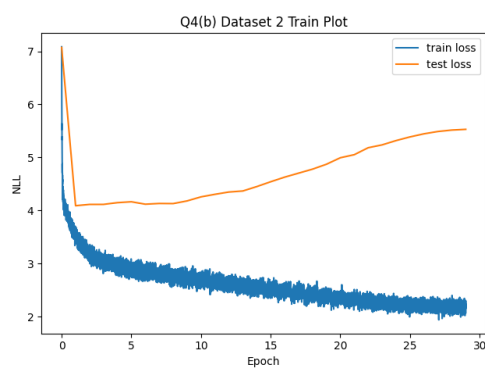


(a) Dataset 1: Training curve



(b) Dataset 1: Samples

Final test loss for dataset 2: **FILL IN HERE** nats / dim



(a) Dataset 2: Training curve



(b) Dataset 2: Samples

Question 5: Causal Transformer - Text

(a) [20pt] Modeling Text

Final test loss: **FILL IN HERE** nats / dim

Test Sample

Test Sample 1

What more is more for what that know. That only is
but and more beautiful than. That is the
best? Of course, as for

Test Sample 2

When things are things, who should have the
perfection? They place their place. For what
which cannot appear to be

Test Sample 3

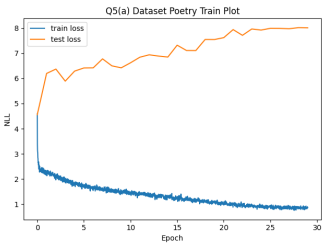
Has been and all. For what is it, to that the
Let there be the first of the first. It is
good to have the first

Test Sample 4

When, should have been the most common case?
There is no reason to be and it is not the most
any more. It is

Test Sample 5

For what is it, to that the first of the first
Let there be the first of the first. It is
good to have the first



(a) Training curve

(b) Text samples

Question 6: Causal Transformer - Multimodal

(a) [20pt] Multimodal Text and Image Generation

Final test loss: **FILL IN HERE** nats / dim