Homework 1: Autoregressive Models

Name: Eli Bronstein
Student ID: 26086997

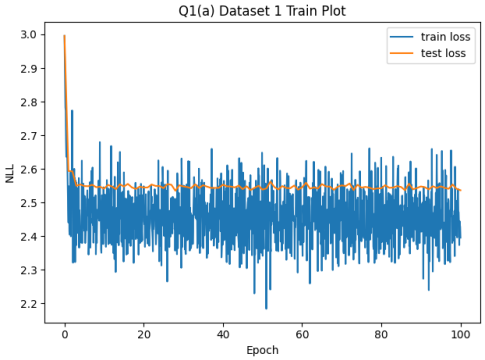
Deliverable: This PDF write-up by **Tuesday February 7th, 23:59pm**. Your PDF should be generated by simply replacing the placeholder images of this LaTeX document with the appropriate solution images that will be generated automatically when solving each question. The solution images are automatically 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!

Note: My submission includes this PDF with all the figures and test losses, the iPython notebook I primarily used for prototyping and experimentation, and the `hw1.py` script, which I ran to get final results for most of the questions (mostly question 3 onwards). The notebook has some of the final figures and test losses, while the rest were produced by running the Python script.

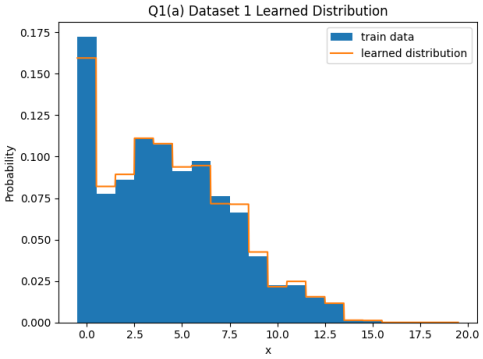
Question 1: 1D Data

(a) [10pt] Fitting a Histogram

Final test loss for dataset 1: 2.5365 nats / dim

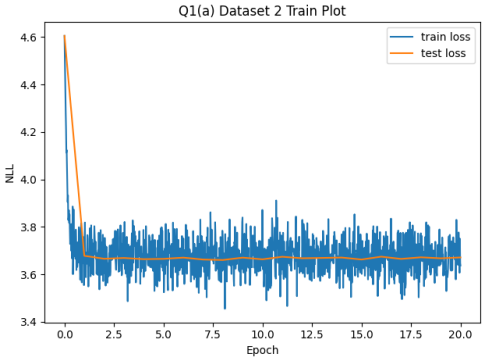


(a) Dataset 1: Training curve

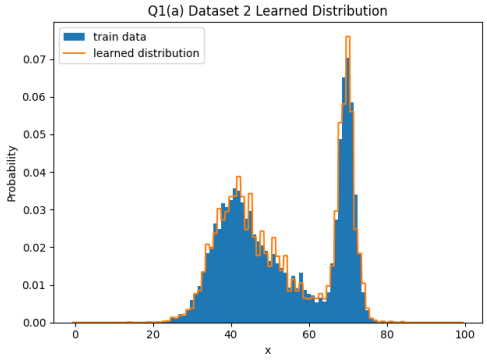


(b) Dataset 1: Learned distribution

Final test loss for dataset 2: 3.6712 nats / dim



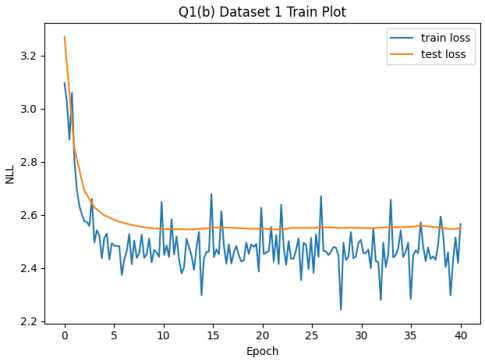
(a) Dataset 2: Training curve



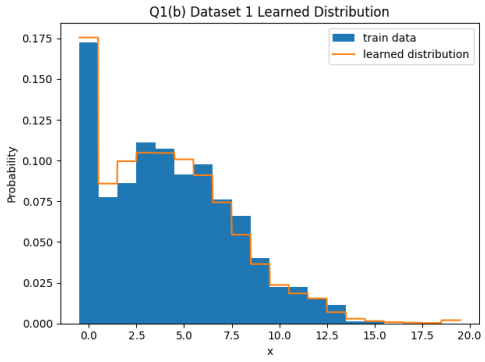
(b) Dataset 2: Learned distribution

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

Final test loss for dataset 1: 2.5495 nats / dim

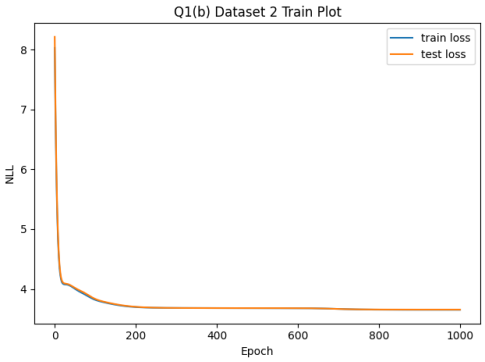


(a) Dataset 1: Training curve

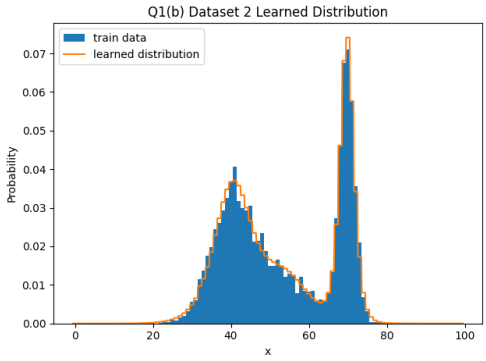


(b) Dataset 1: Learned distribution

Final test loss for dataset 2: 3.6542 nats / dim



(a) Dataset 2: Training curve

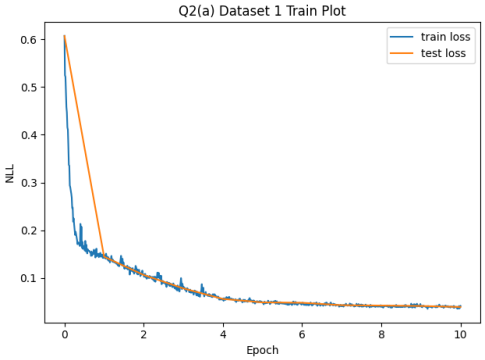


(b) Dataset 2: Learned distribution

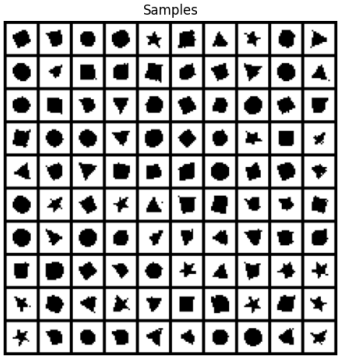
Question 2: PixelCNNs

(a) [15pt] PixelCNNs on Shapes and MNIST

Final test loss for dataset 1: 0.0396 nats / dim

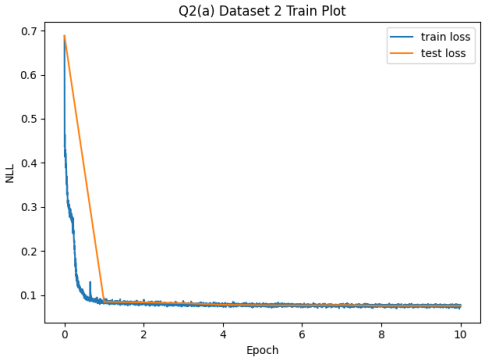


(a) Dataset 1: Training curve

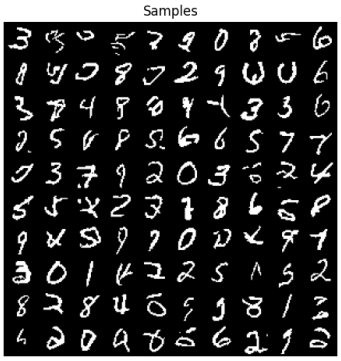


(b) Dataset 1: Samples

Final test loss for dataset 2: 0.0757 nats / dim



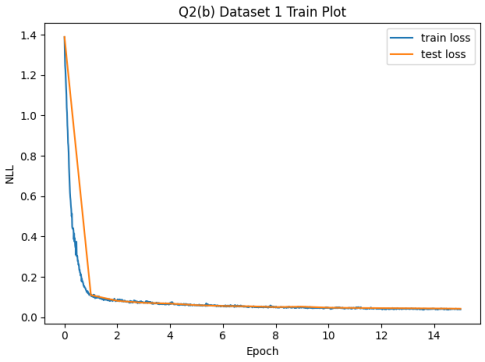
(a) Dataset 2: Training curve



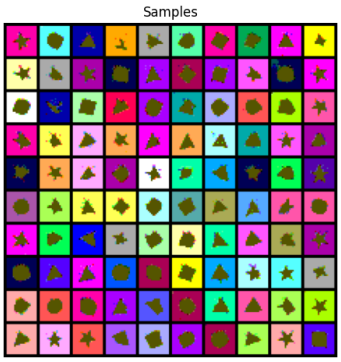
(b) Dataset 2: Samples

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

Final test loss for dataset 1: 0.0425 nats / dim

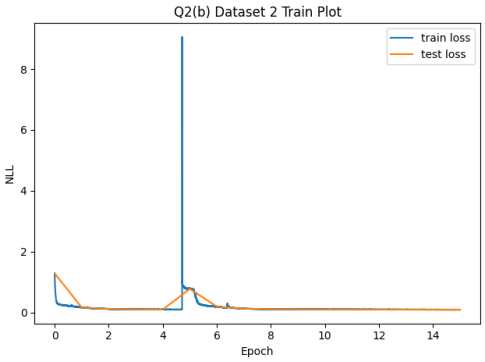


(a) Dataset 1: Training curve



(b) Dataset 1: Samples

Final test loss for dataset 2: 0.0892 nats / dim



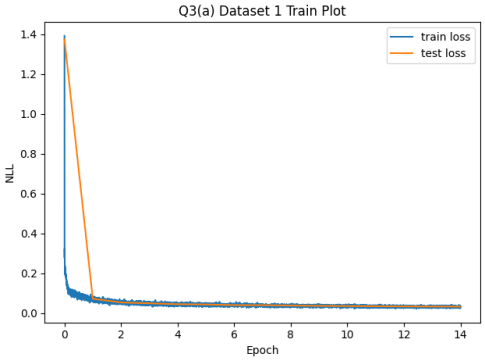
(a) Dataset 2: Training curve

(b) Dataset 2: Samples

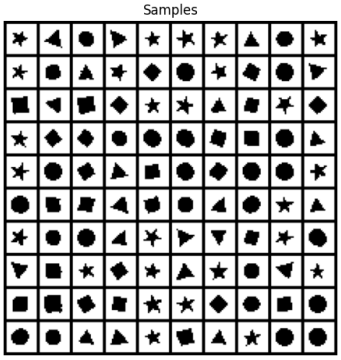
Question 3: Causal Transformer - iGPT

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

Final test loss for dataset 1: 0.0337 nats / dim

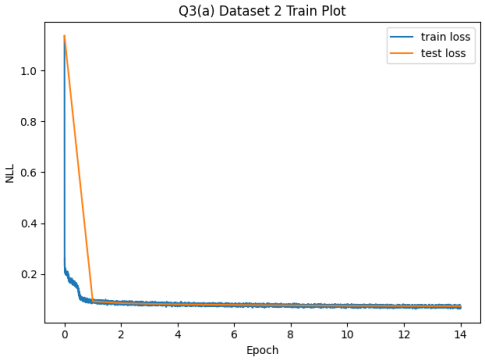


(a) Dataset 1: Training curve

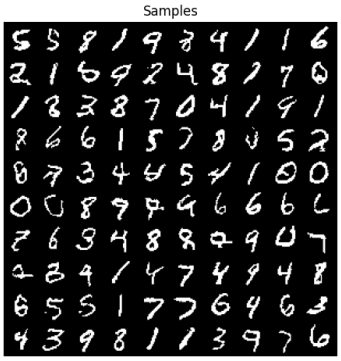


(b) Dataset 1: Samples

Final test loss for dataset 2: 0.0741 nats / dim



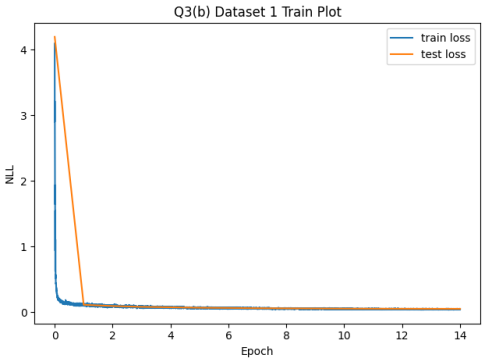
(a) Dataset 2: Training curve



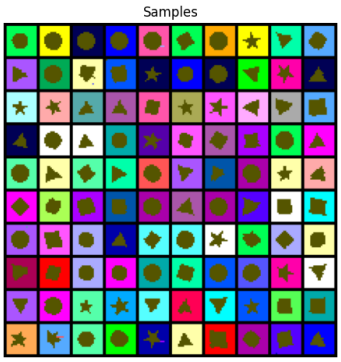
(b) Dataset 2: Samples

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

Final test loss for dataset 1: 0.0503 nats / dim

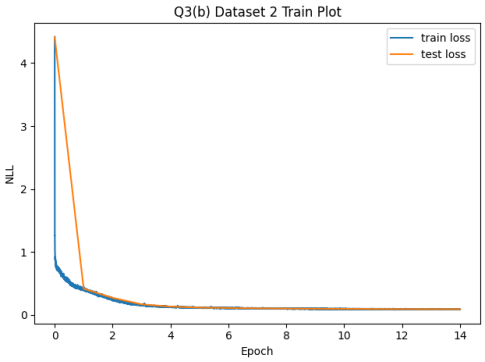


(a) Dataset 1: Training curve



(b) Dataset 1: Samples

Final test loss for dataset 2: 0.0918 nats / dim



(a) Dataset 2: Training curve

(b) Dataset 2: Samples

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

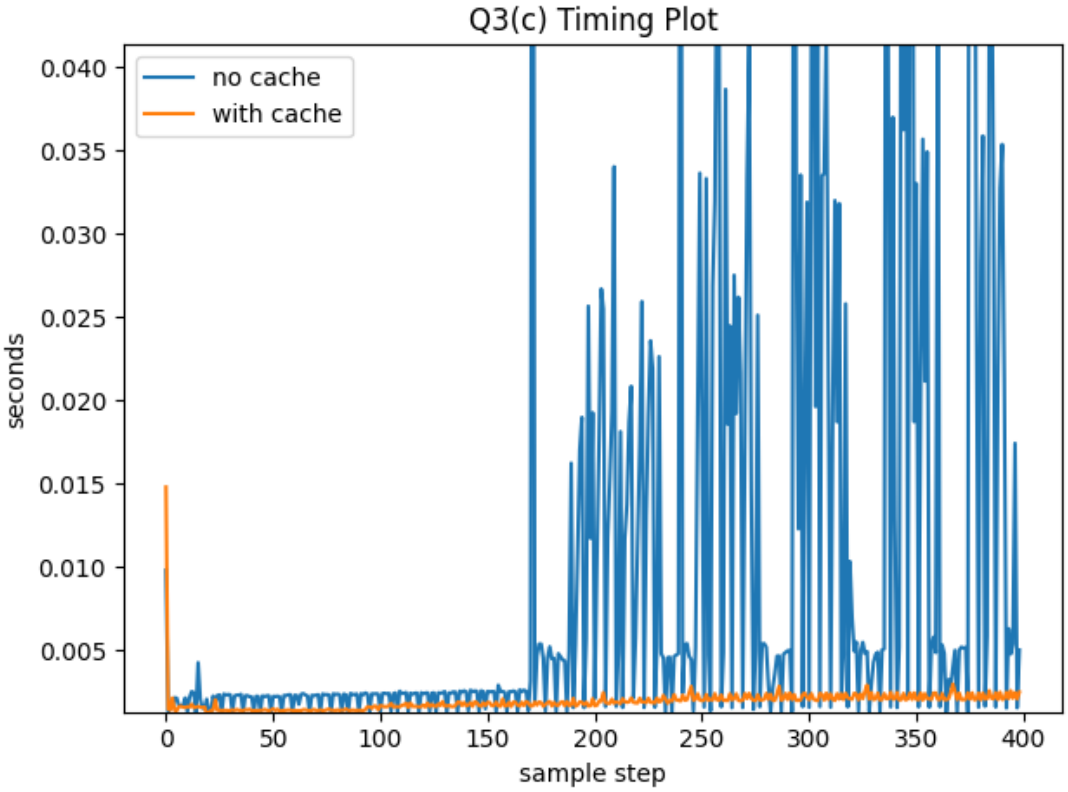
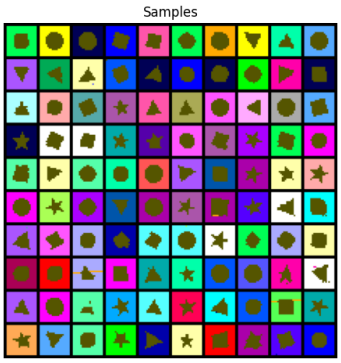
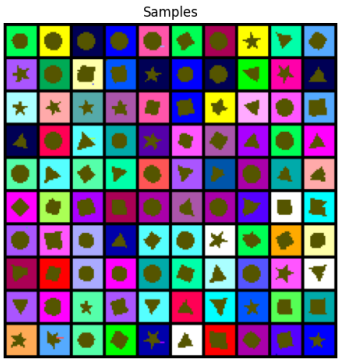


Figure 13: Dataset 2: Inference Speed



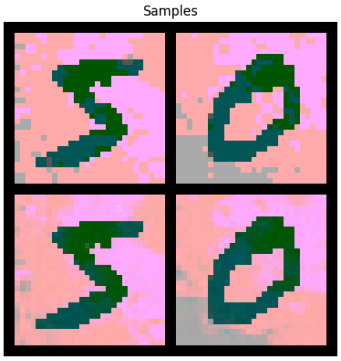
(a) Dataset 2: Samples (no caching)



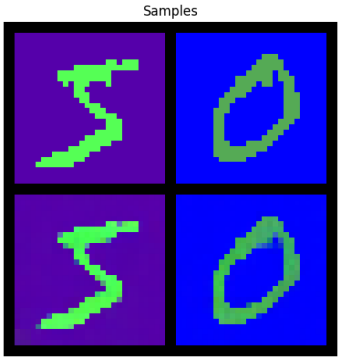
(b) Dataset 2: Samples (caching)

Question 4: Causal Transformer - Tokenized Images

(a) **[5pt] Image Quantization**



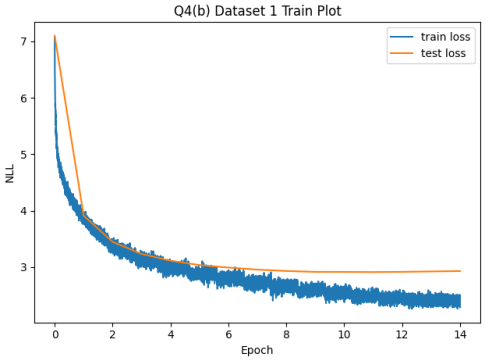
(a) Dataset 1: Quantized Examples



(b) Dataset 2: Quantized Examples

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

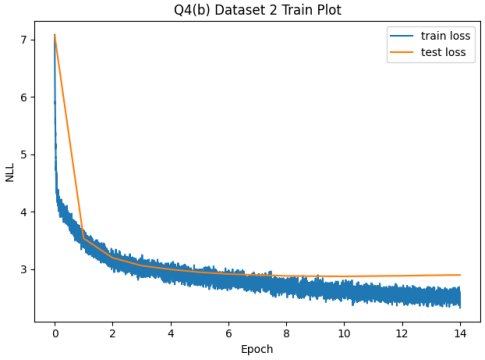
Final test loss for dataset 1: 2.9278 nats / dim



(a) Dataset 1: Training curve

(b) Dataset 1: Samples

Final test loss for dataset 2: 2.9000 nats / dim



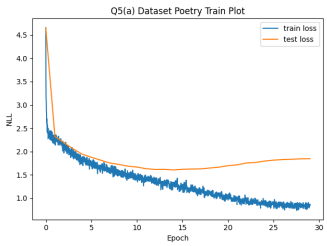
(a) Dataset 2: Training curve

(b) Dataset 2: Samples

Question 5: Causal Transformer - Text

(a) **[20pt] Modeling Text**

Final test loss: 1.8461 nats / dim



(a) Training curve

Text Samples

Text Sample 1
Alas, I come, I my hand all my love.

Text Sample 2
Orphen S Its it his deeine is to his eye I see mortalitous of one any control When Ocean's have longing sorrow of forbearth.

Text Sample 3
Farewed content, many now elpeast, Adieugh thine all table bridanches are renneed, And think wisher'd wayrds checeefered oft

Text Sample 4
Shy ender still, sike she worthnd a rooth waste With blessed companious, silves, if Ill love Not o'clockbird in win

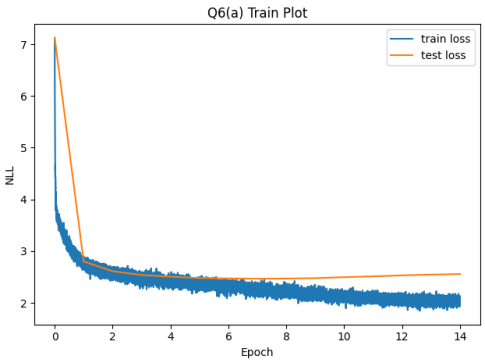
Text Sample 5
Thou art my winter? I must inferacts their man, But plood always not time to sing, for they fight Slow I such fair watery and

(b) Text samples

Question 6: Causal Transformer - Multimodal

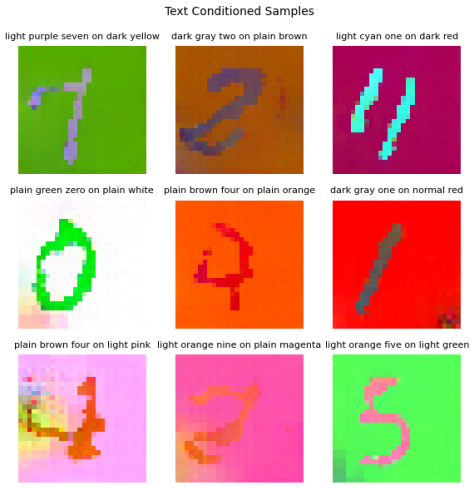
(a) [20pt] Multimodal Text and Image Generation

Final test loss: 2.5597 nats / dim

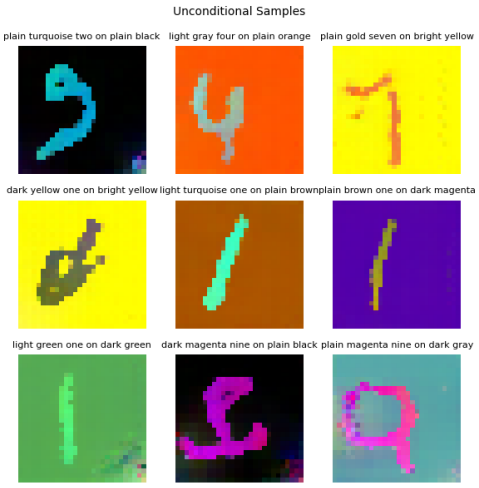


(a) Training curve

(b) Image conditioned samples



(c) Text conditioned samples



(d) Unconditional samples