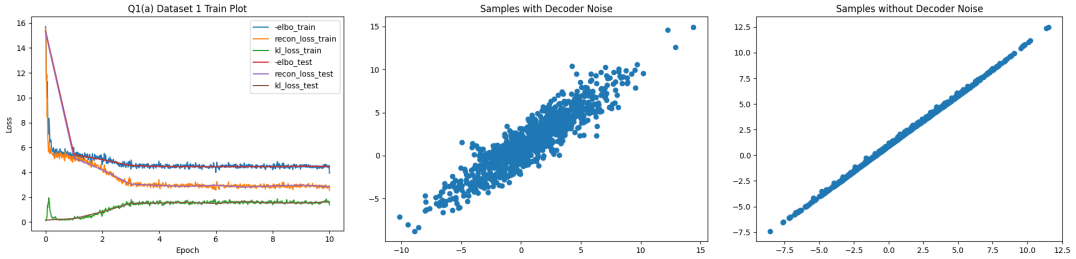
Homework 2: Latent Variable Models

Deliverable: This PDF write-up by **Wed February 21st, 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.

Question 1: VAEs on 2D Data [20pt]

(a) [10pt] Data from a Full Covariance Gaussian

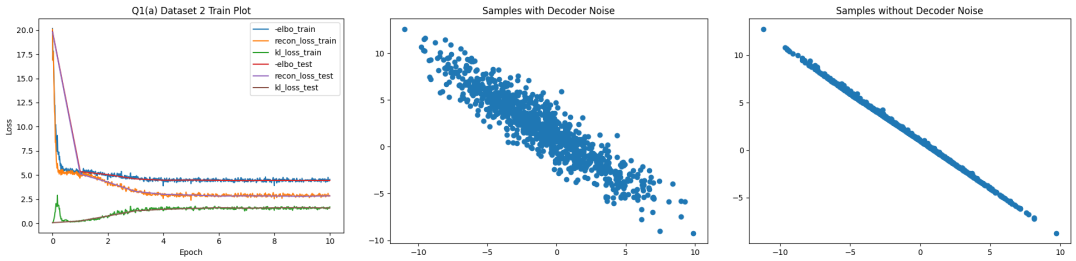
Final Full -ELBO: 4.4710, Recon Loss: 2.8516, KL Loss: 1.6194 (Dataset 1)



(a) Training curve (b) Samples with Noise (c) Samples without Noise

Figure 1: Results for Dataset 1

Final Full -ELBO: 4.4429, Recon Loss: 2.8461, KL Loss: 1.5967 (Dataset 2)

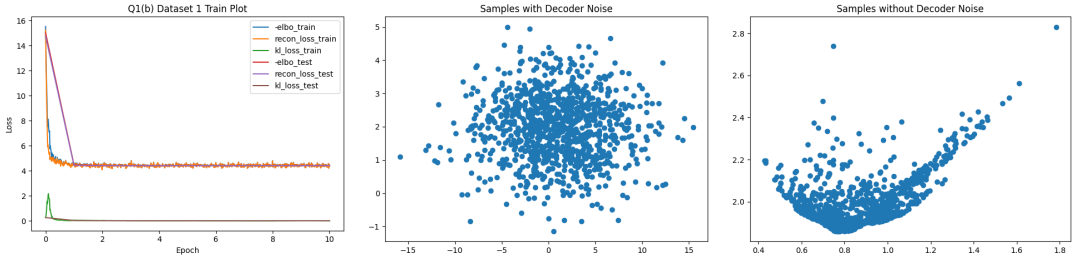


(a) Training curve (b) Samples with Noise (c) Samples without Noise

Figure 2: Results for Dataset 2

(b) **[10pt] Data from a Diagonal Gaussian**

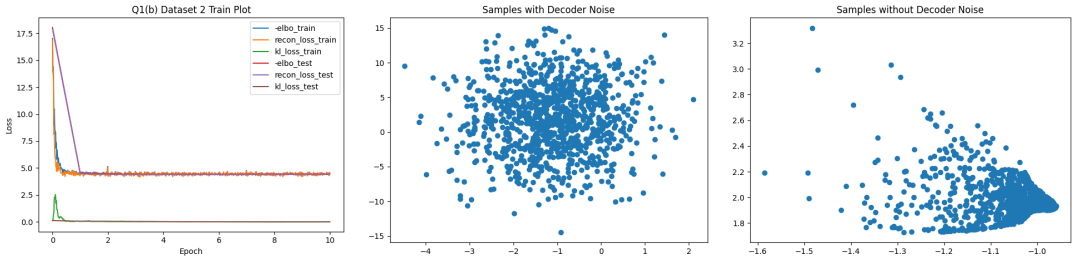
Final Full -ELBO: 4.4237, Recon Loss: 4.4119, KL Loss: 0.0118 (Dataset 1)



(a) Training curve (b) Samples with Noise (c) Samples without Noise

Figure 3: Results for Dataset 1

Final Full -ELBO: 4.4303, Recon Loss: 4.4206, KL Loss: 0.0097 (Dataset 2)



(a) Training curve (b) Samples with Noise (c) Samples without Noise

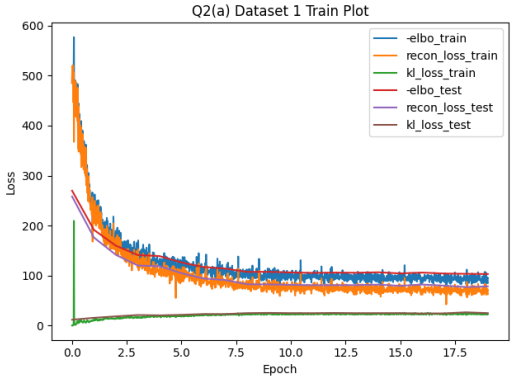
Figure 4: Results for Dataset 2

Answer: The latents are being used for the datasets in part (a) but not in part (b). Since the data in part(b) comes from a Normal distribution with a diagonal covariate matrix, both the latents and the data are 2D, and latent prior is also a diagonal Gaussian, the latents are not needed to recover the data distribution. The latents can simply be mostly constant, which can be seen from the no-noise samples in part (b) that are mostly clustered around $[-1, 2]$ (i.e., near the mean of the data). The learned decoder variance and the samples $\epsilon \sim \mathcal{N}(0, 1)$ are sufficient to accurately describe the data. In contrast, the dataset in part (b) has a full covariance matrix, so using a constant latent plus a learned covariance times the noise is insufficient to represent it. For this reason, we see that the samples without noise actually look like a de-noised version of the dataset.

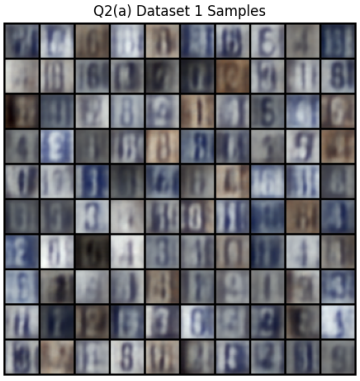
Question 2: VAEs on Images [40pt]

(a) **[20pt] VAE**

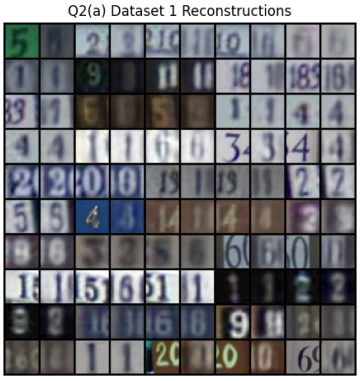
Final Full -ELBO: 102.9934, Recon Loss: 78.2405, KL Loss: 24.7529 (Dataset 1)



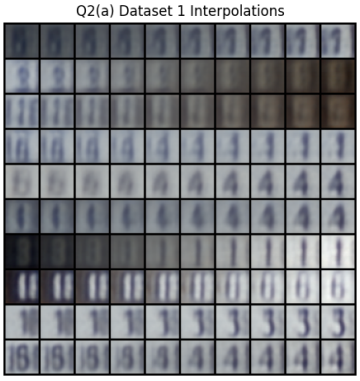
(a) Training Curve



(b) Samples



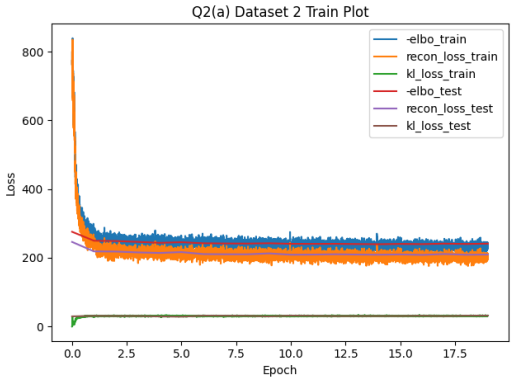
(c) Reconstructions



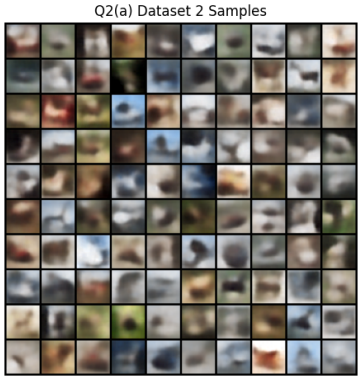
(d) Interpolations

Figure 5: Results for Dataset 1

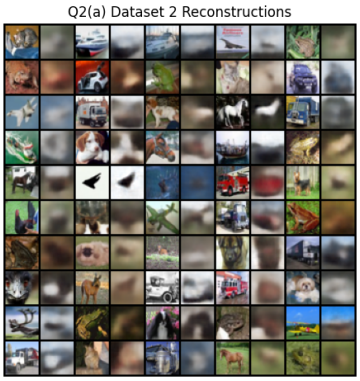
Final Full -ELBO: 241.2836, Recon Loss: 209.5040, KL Loss: 31.7796 (Dataset 2)



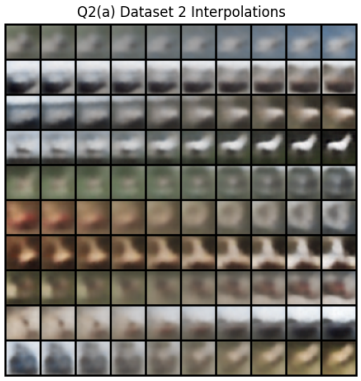
(a) Training Curve



(b) Samples



(c) Reconstructions

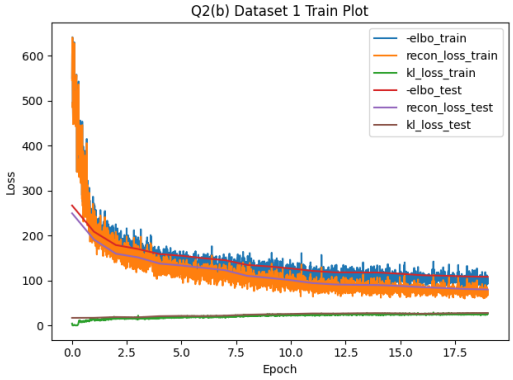


(d) Interpolations

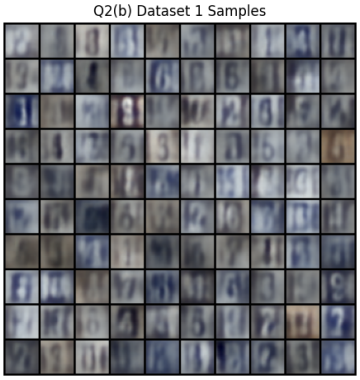
Figure 6: Results for Dataset 2

(b) **[20pt] Hierarchical VAE with Learned Prior**

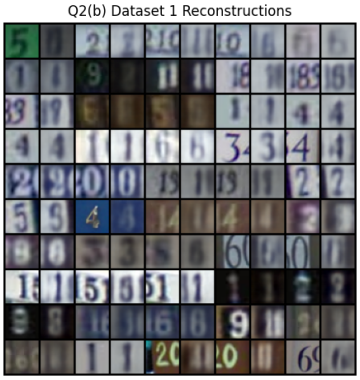
Final Full -ELBO: 108.6899, Recon Loss: 80.4940, KL Loss: 28.1959



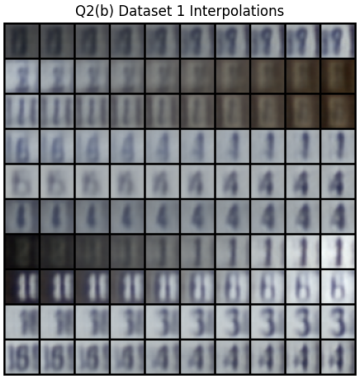
(a) Training Curve



(b) Samples



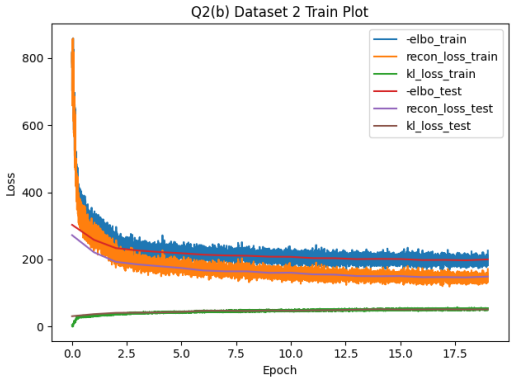
(c) Reconstructions



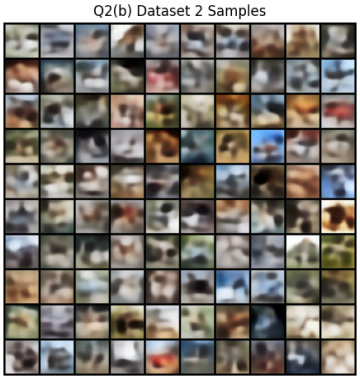
(d) Interpolations

Figure 7: Results for Dataset 1

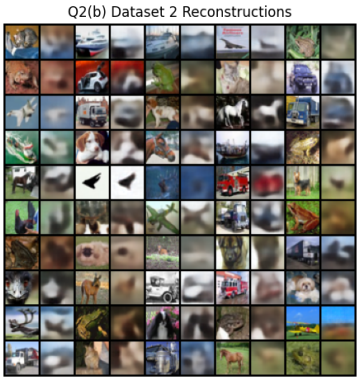
Final Full -ELBO: 200.0667, Recon Loss: 148.9151, KL Loss: 51.1517 (Dataset 2)



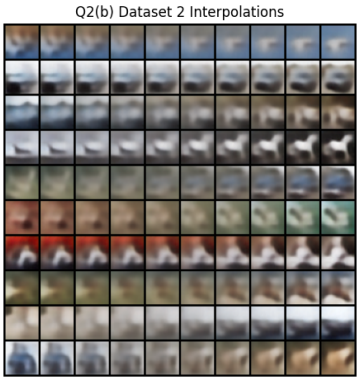
(a) Training Curve



(b) Samples



(c) Reconstructions

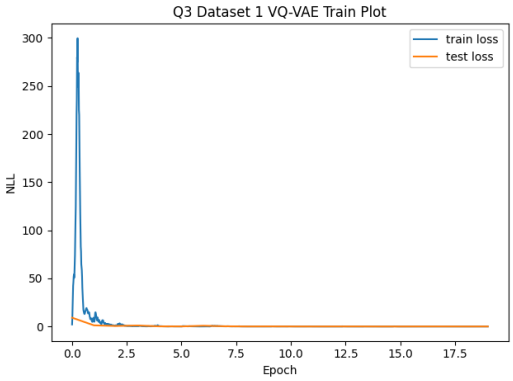


(d) Interpolations

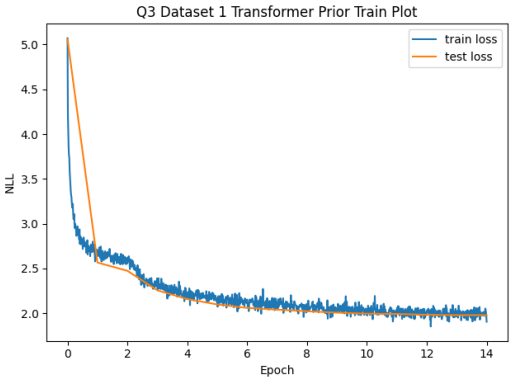
Figure 8: Results for Dataset 2

Question 3: VQ-VAE [40pt]

Final VQ-VAE Test Loss: 0.0630, Transformer Prior Test Loss: 1.9789 (Dataset 1)

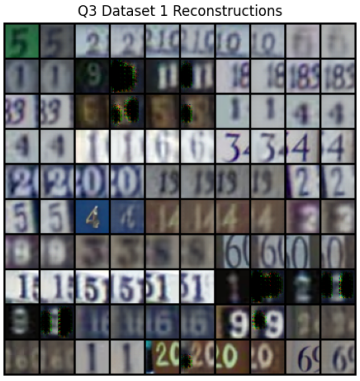


(a) VQ-VAE Training Curve



(b) Transformer Prior Training Curve

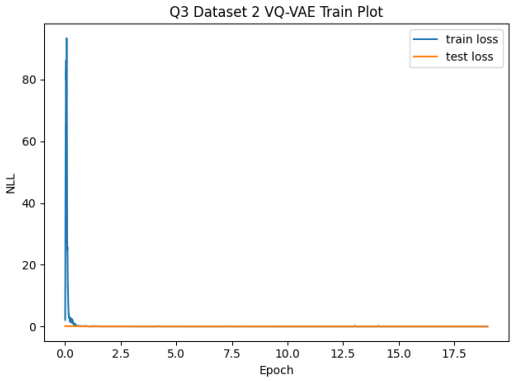
(c) Samples



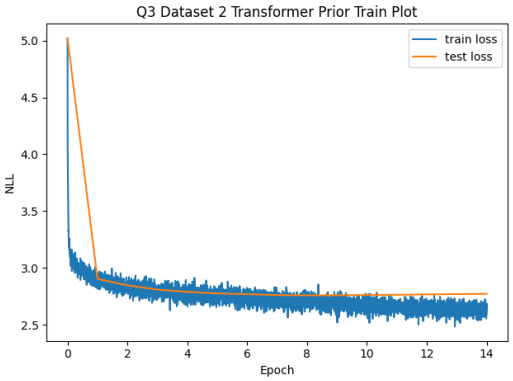
(d) Reconstructions

Figure 9: Results for Dataset 1

Final VQ-VAE Test Loss:0.0451, Transformer Prior Test Loss: 2.7739 (Dataset 2)



(a) VQ-VAE Training Curve



(b) Transformer Prior Training Curve

(c) Samples

(d) Reconstructions

Figure 10: Results for Dataset 1