Task 4: Variational Autoencoder (VAE)

Variational Auto Encoders learns probabilistic gaussian latent distribution of data. The encoder maps images to the latent space. The decoder reconstructs images from the latent distribution. Learning is made possible through the use of the reparameterization trick. I have explored different techniques to reconstruct the best representation of images. This includes exploring active z-dimensions, role of KL and reconstruction loss on regenerations and sampling from different distributions i.e. normal or laplacian.

1. Train the VAE

I have trained a VAE on Fashion-MNIST with a Gaussian decoder (MSE) and a standard Normal prior. Loss curves show stable optimization: reconstruction error fell from 28.63 to 14.88 MSE per sample while the KL term rose from 6.13 and plateaued at approx. 7.98 nats, yielding a total ELBO around 22.86 as in Fig. (4.1)

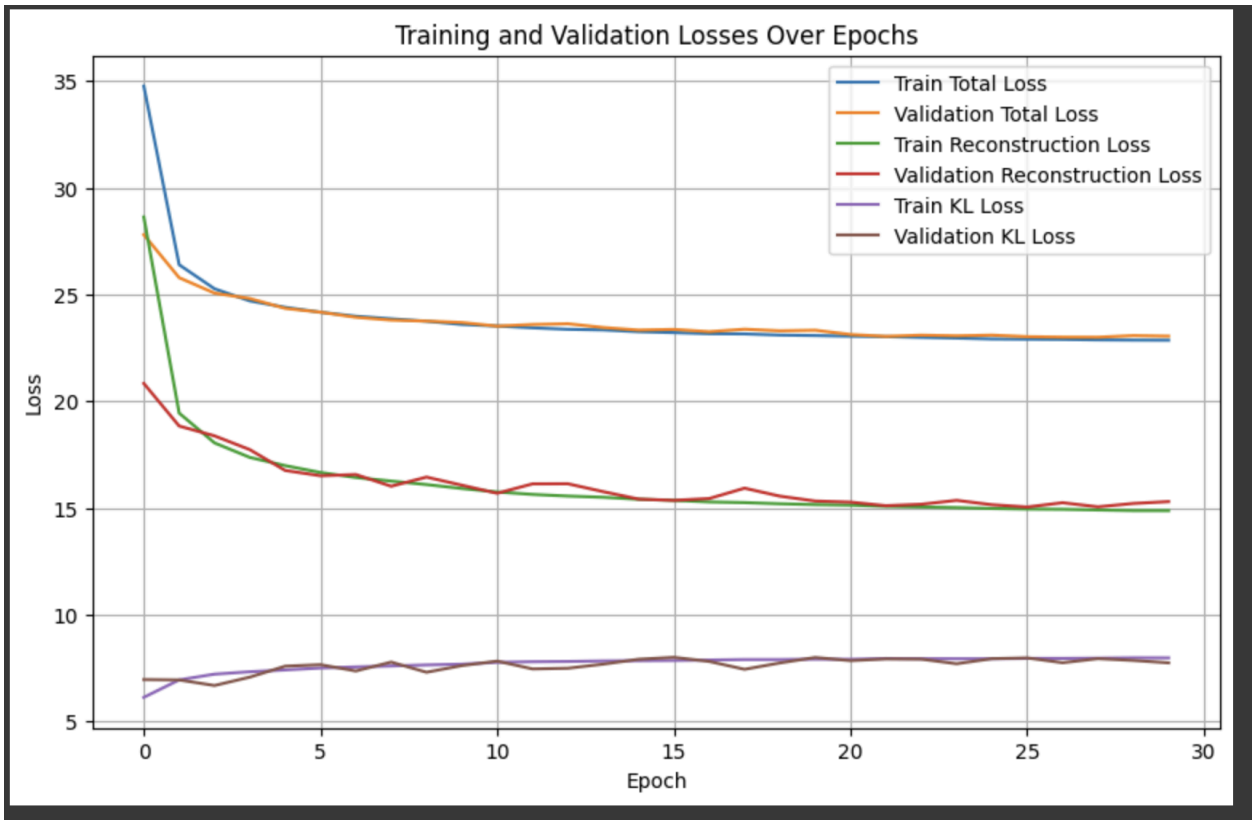


Fig (4.1)

2. Visualize Reconstructions and Generations

Reconstructions are compared with originals and it was observed that reconstructions are blurry and only preserves edges and silhouettes. We can see that textures and texts are all gone as Gaussian approximates pixels.

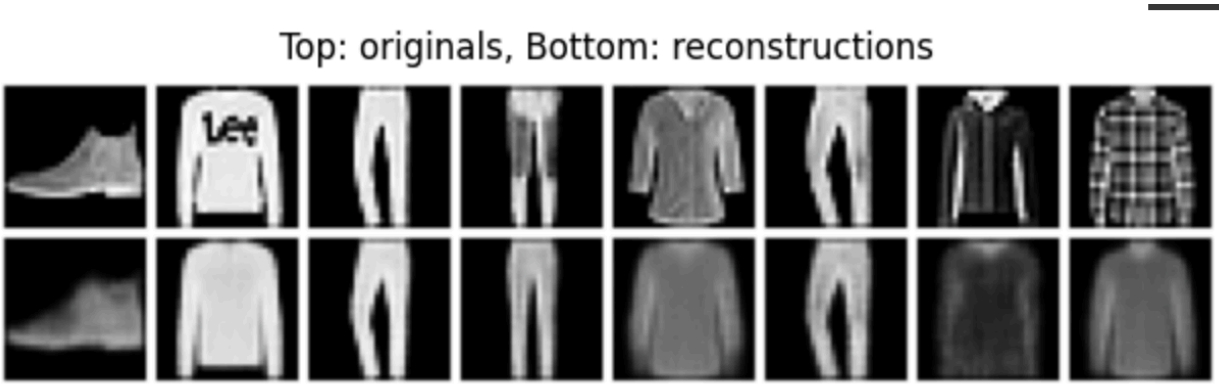


Fig (4.2)

3. Gaussian vs laplacian generations

Sampling from the Normal prior has produced low quality images, the reason could be a small number of epochs and because most dimensions were unused and only 6 out of 20 dimensions have contributed to the results. The encoder did not have enough to learn, sampling from a Laplace prior also degrades quality due to prior-decoder mismatch. Since images are reconstructed from Gaussian latent space of 20dim, the images are blurry and edges aren’t refined.

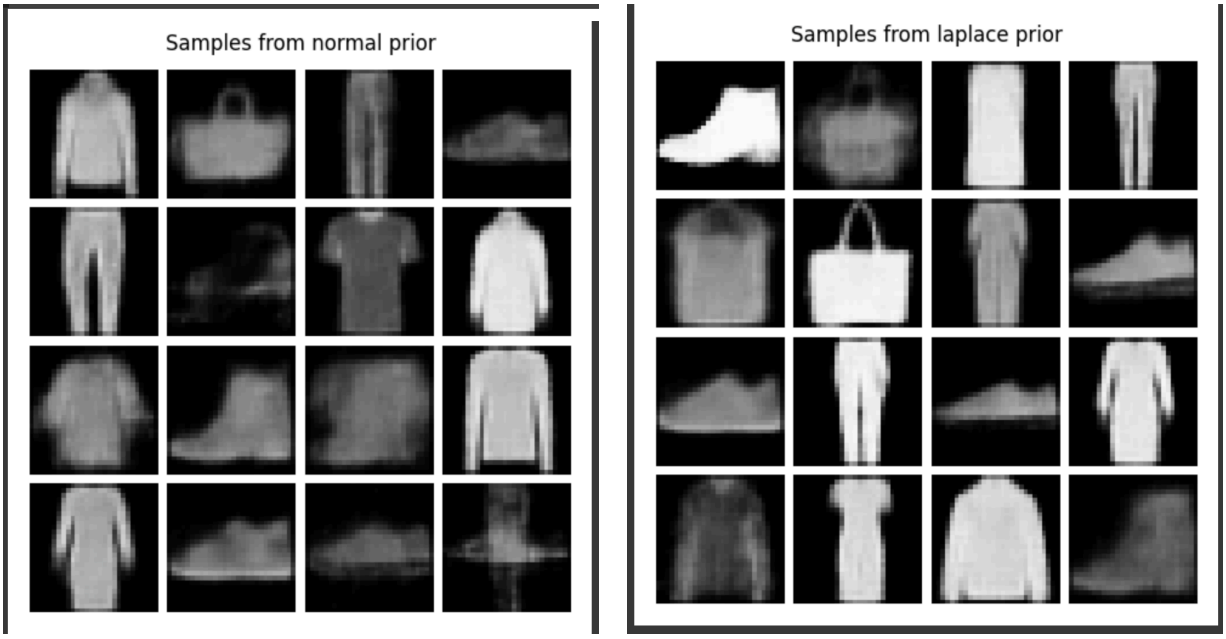
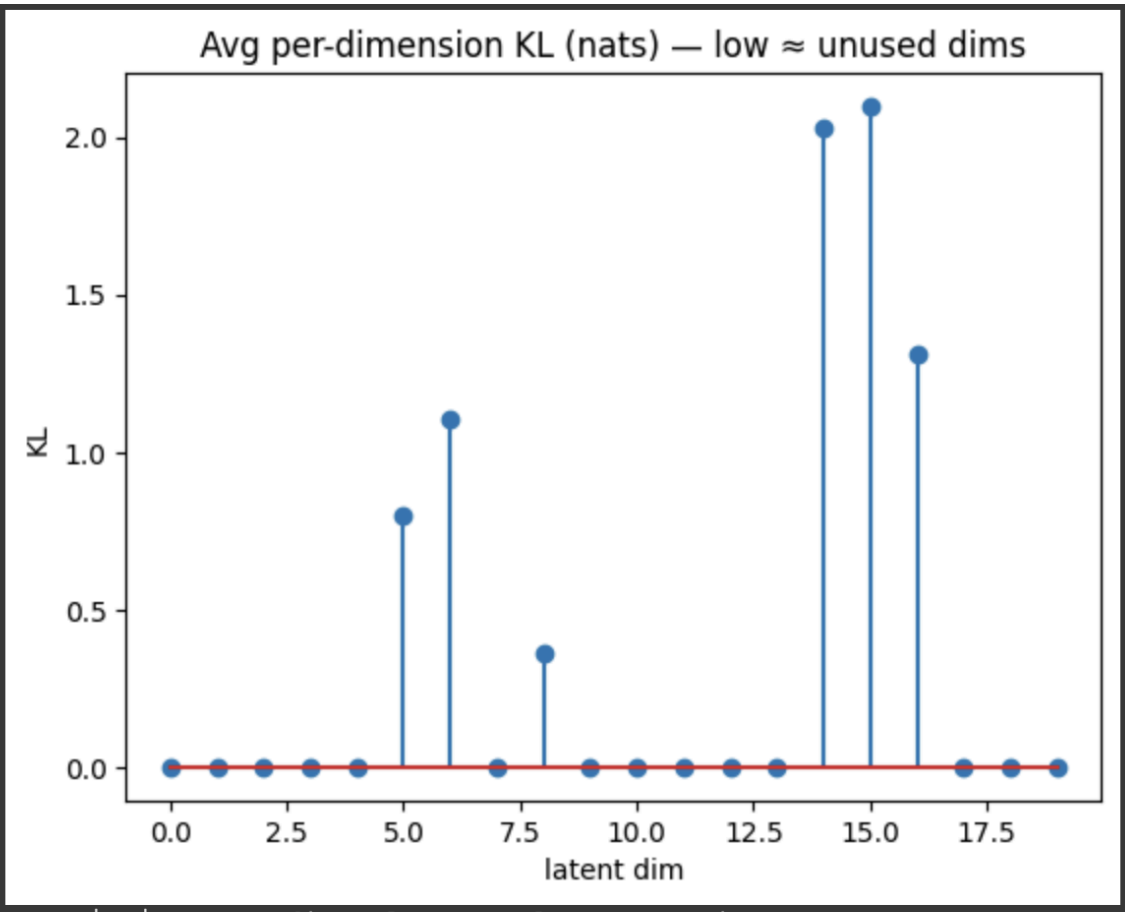


Fig (4.3)

4. Posterior Collapse

Plot of Fig (4.2) clearly shows which latent coordinates are used by model. For most dimensions KL is zero and only 6 dimensions show a spike with considerable KL values. The stats show that mean per dimension is small and log per variance is highly negative for the active dimensions. From the stats we can see that KL is largely variance-driven.




```
Mean |mu| across dims (lower ≈ closer to prior mean):
[0.01753526 0.04362604 0.0040834  0.0031585  0.02009211 0.07844752
 0.03521003 0.02037184 0.06009704 0.02184301 0.00939641 0.02512023
 0.01715139 0.03444309 0.07379159 0.04893012 0.0531049  0.01131726
 0.00419823 0.02459726]
Mean logvar across dims (≈0 means var ~ 1):
[-2.0975057e-02 -1.3847530e-02 -5.6832889e-03 -7.7015832e-03
 -2.6408961e-02 -1.6777885e+00 -2.2694016e+00 -2.0041941e-03
 -7.3168087e-01  2.5753400e-03 -2.2047838e-02 -2.9426817e-02
 -8.5356021e-03 -2.3386558e-03 -4.1056337e+00 -4.1835051e+00
 -2.6950746e+00 -7.3109358e-04 -4.6359818e-02 -6.1994134e-03]
Total KL per sample (approx): 7.724 nats
```

Fig(4.4)

4. Mitigating Posterior Collapse

To mitigate the effect of posterior collapse techniques of warm-up = 10 and free bits = 0.05 are introduced. This was to give more room for the encoder to learn and map images in latent space. Resultantly images reconstructed have better quality and instead of 4 dimensions now all 20 dimensions are used.

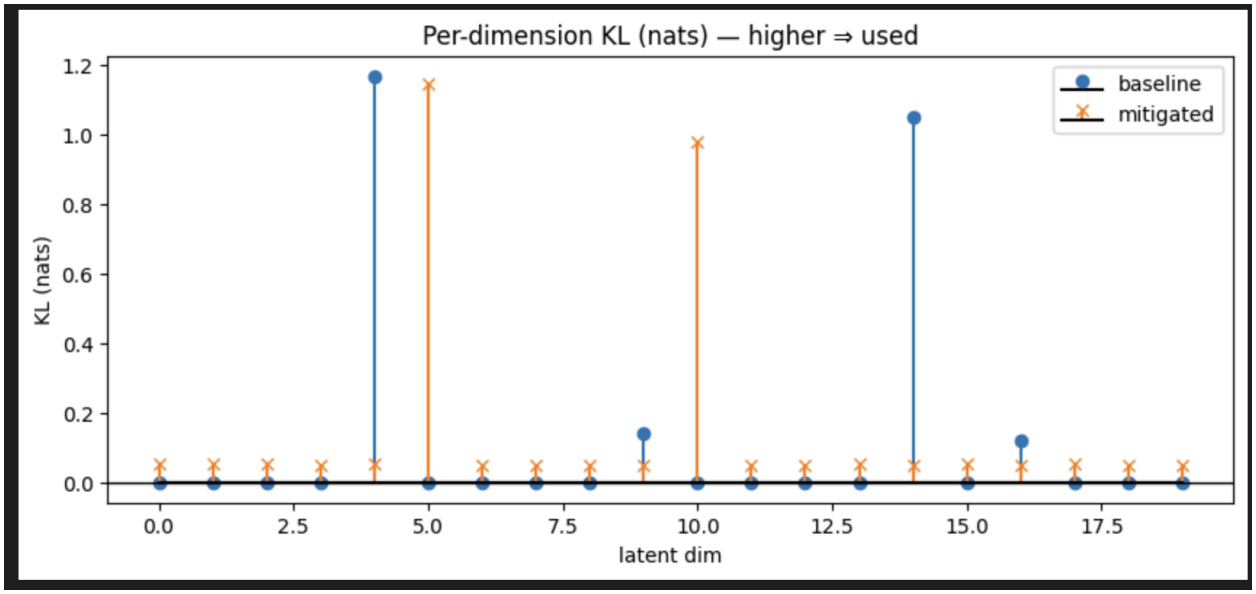

Per-dimension KL (nats) — higher ⇒ used
baseline
mitigated
KL (nats)
latent dim
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

Fig (4.5)

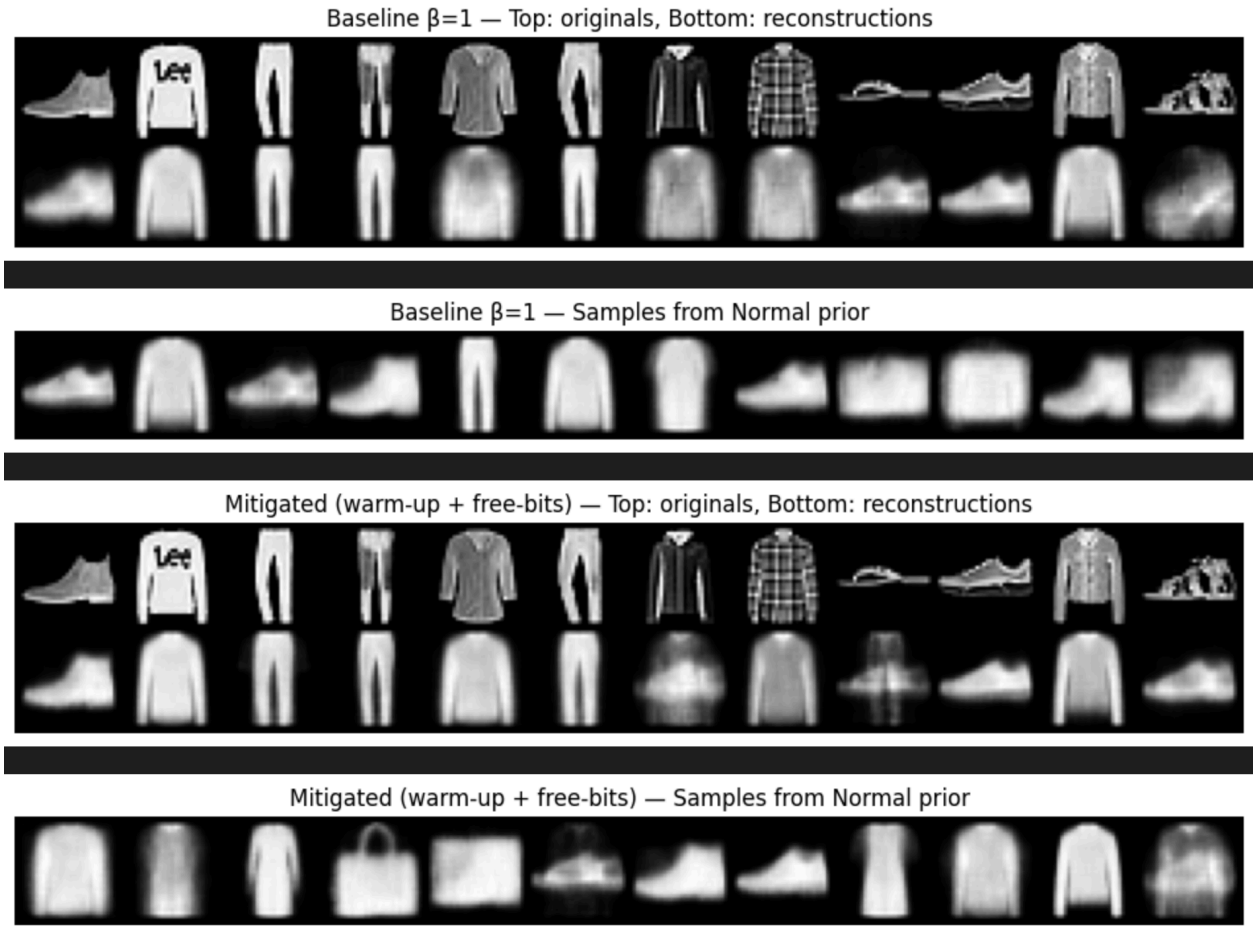

Baseline β=1 — Top: originals, Bottom: reconstructions
Baseline β=1 — Samples from Normal prior
Mitigated (warm-up + free-bits) — Top: originals, Bottom: reconstructions
Mitigated (warm-up + free-bits) — Samples from Normal prior

Fig (4.6)

Conclusion:

Per-dimension analysis reveals partial posterior collapse, despite a 20-D latent, only 5–6 dimensions carry substantial KL, largely via variance shrinkage (means near zero, strongly negative log-variances on active dims), therefore reconstructions preserve category and structure but lose fine details.WarmUp and free bits were introduced to enhance the quality of reconstructed images. Reconstruction is consistent with the information budget and the Gaussian objective. Overall, the model learns a compact, smooth latent representation that supports plausible generations and reconstructions.

Some other possible ways to improve the model can be:

1) Change of objective function, we can try using BCEwithlogits on FashionMNIST dataset as Bernoulli can preserve edges and details better than MSE
2) Increasing dimensions of latent space. It may increase KL but can reduce reconstruction error