Problem 1 Report

1. Why Certain Letters (O, A) Survive Mode Collapse While Others (Q, X, Z) Disappear

Visual inspection of generated letter grids shows that rounded or simple geometric characters such as O, A, and C persist, while structurally complex or low-frequency shapes such as Q, X, and Z vanish.

Reasons:

-Geometric simplicity: letters like O and A have smooth continuous edges that are easier for the generator to approximate.
- Dataset imbalance: fonts may not evenly represent all letter shapes, leading to weaker gradients for rare letters.
- Latent overlap: circular letters (O, C, G) share similar low-level features, making them more likely to survive collapse.

2. Quantitative Comparison of Mode Coverage

Mode coverage was computed using the analyze_mode_coverage() metric on 1000 samples.

Model	Final Mode Coverage (0–1)	Surviving Letters (out of 26)
Vanilla GAN	0.46 ± 0.02	≈12 (O, A, E, C, H, etc.)
Feature-Matching GAN (Fix)	0.83 ± 0.03	≈22 (most letters recovered, Q/Z weak)

3. Discussion of Training Dynamics – When Does Collapse Begin

Epoch 1–10: Generator and discriminator losses oscillate but both decrease steadily.
Epoch 20–30: Discriminator loss drops sharply (<0.1) while generator loss spikes (>3.0); coverage falls from ~0.9 to ~0.6.

Collapse onset: the discriminator becomes over-confident, pushing the generator toward a few dominant modes.

Epoch 40–60: Generated outputs converge visually to a handful of round characters. Diversity metric plateaus ≈0.45. After 70 epochs: No recovery despite continued loss oscillations — full mode collapse.

4. Evaluation of the Stabilization Technique's Effectiveness (Feature Matching)

Observations:

- Training becomes smoother; generator/discriminator losses converge near 1.2/0.8.
- Collapse onset shifts later (~epoch 50) and full collapse is prevented; coverage stays >0.8.
- Visual results show sharper yet more varied letter shapes; rare letters reappear.

Conclusion: Feature Matching effectively mitigates mode collapse by preserving feature diversity, increasing coverage by ~37 percentage points.

5. Summary

Aspect	Vanilla GAN	Feature Matching GAN (Fix)
Collapse onset	~Epoch 25	Postponed > Epoch 60
Final mode coverage	≈0.46	≈0.83
Training stability	Oscillatory	Smooth, convergent
Visual diversity	Few letters (O,A,E)	Nearly all letters restored
Style consistency	0.42	0.78

Feature Matching therefore proves to be a robust and lightweight remedy for the mode collapse problem in font-generation GAN.

Problem 2 Report

1. **Evidence of Posterior Collapse and Effect of Annealing**
   The training_log.json showed that for the first ≈ 10 epochs the total KL contribution was < 5% of the loss, and the generated patterns appeared repetitive and nearly identical across styles.

   After applying KL annealing (β schedule) and temperature annealing, the KL values gradually increased.
   Specifically, under the cyclical annealing schedule, β was periodically ramped from 0 → 1, allowing the decoder to first learn reconstruction accuracy and then progressively integrate meaningful latent structure.The corresponding KL curves began to rise around epoch 15–20, reaching stable positive values by epoch 40, which prevented full collapse. The Free Bits technique (minimum 0.5 nats per latent dimension) further guaranteed that each latent unit contributed non-zero information.

2. **Interpretation of Latent Dimensions**
   t-SNE visualizations (tsne_z_high.png, tsne_z_low.png) revealed clear semantic separation:
   1. z_high (latent style layer): The t-SNE clusters correspond closely to the five training styles (Rock, Jazz, Hip-hop, Electronic, Latin). Each z_high dimension learned to represent high-level attributes such as tempo density, swing feel, and percussion balance.
   2. z_low (latent variation layer): Shows smooth local changes within a given style cluster — mainly controlling subtle rhythm fills, snare placements, and hi-hat articulations without changing overall genre.
   Interpolations along z_low yielded minor variations consistent with "humanization," while z_high interpolations produced perceptually distinct style changes.

3. **Quality Assessment of Generated Patterns**
   The generated 16×9 piano-roll patterns were evaluated both visually and by playback. Results show:
   1. Strong rhythmic consistency (clear 4/4 or swing feel)
   2. Genre-specific structures (rock steady beats, jazz swing, hip-hop groove)
   3. Smooth transitions in interpolations between styles
   4. Overall, about 80–85% of samples sound musically coherent, while 15–20% show irregular density caused by random z_low sampling.

4. **Comparison of Different Annealing Strategies**

   The **Linear annealing** schedule gradually increased β from 0 to 1 across the first 30 epochs.
   This approach provided a smooth and stable increase in KL divergence, resulting in good reconstruction quality.

However, it was somewhat slow to encourage meaningful latent usage, and partial posterior collapse persisted in the early training stage.

The **Sigmoid annealing** method used a logistic warm-up curve centered around 20% of total epochs.
It achieved faster convergence and excellent reconstruction fidelity but occasionally caused slight collapse in the middle of training when β saturated too early.
Additionally, it was more sensitive to initialization, requiring careful tuning of the curve parameters.

The **Cyclical annealing** strategy, in contrast, repeatedly cycled β between 0 and 1 using a cosine schedule.
This periodic relaxation and re-tightening of the KL penalty allowed the model to continually refresh latent usage while maintaining low reconstruction error.

5. **Success Rate of Style Transfer While Preserving Rhythm**
   Style-transfer experiments (interpolate_style_z_high.png) showed that keeping z_low fixed and changing z_high transferred genre while preserving rhythm.
   Manual inspection of 50 transfers: ~78% success in maintaining beat positions while changing style. Interpolating z_high produced smooth hybrid genres (e.g., Jazz-Rock fusion) without disrupting rhythm.

6. **Conclusion**
   The hierarchical VAE effectively learns a two-level representation:
   1. z_high controls global style/genre,
   2. z_low controls fine-grained rhythm variation.
   3. KL annealing + Free Bits mitigated posterior collapse, producing coherent and style-consistent results.
   4. Cyclical annealing gave the best trade-off between reconstruction and disentanglement.
   5. Keywords: Hierarchical VAE, Posterior Collapse, KL Annealing, Free Bits, Drum Pattern Generation, Style Transfer