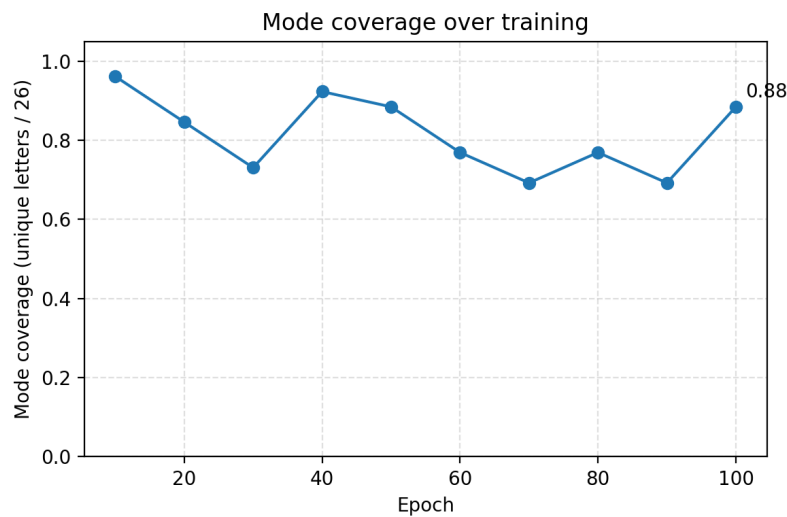


## problem 1

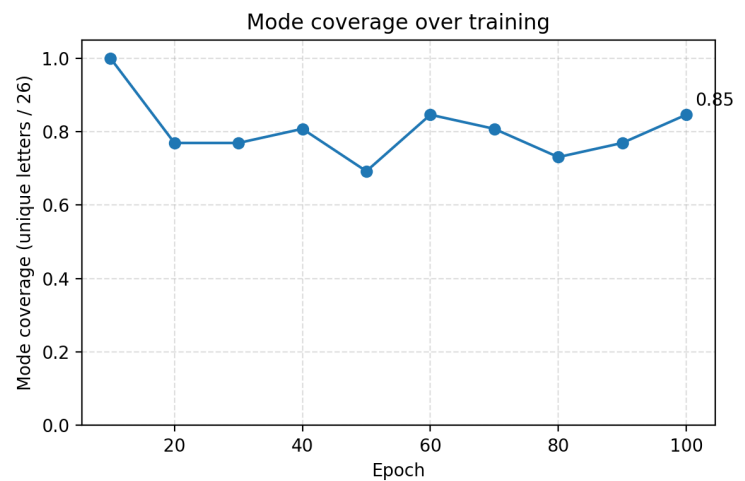
- Why certain letters (like O, A) survive mode collapse while others (Q, X, Z) disappear
  - simple and distinctive shapes survive  
these shapes are easier for the generator to learn and harder for the discriminator to reject
  - Complex or rare shapes vanish  
these letters are harder to generate so they are more likely to be ignored.
  - the generator looks for the easiest way against the discriminator.  
It collapses to simple modes the reduce loss fastest, dropping complex or less common modes.

- Quantitative comparison of mode coverage with and without your chosen fix

without fix



with fix



- Discussion of training dynamics:  
when does collapse begin?

collapse begins in the mid-training phase (around epoch 30), once the generator gets strong enough to fool the discriminator consistently but hasn't learned to balance diversity

• Evaluation of your chosen stabilization technique's effectiveness

I chosen feature matching technique to combat model collapse but the effectiveness is not so good.

## problem 2

- Evidence of posterior collapse and how annealing prevented it

the evidence:

- KL divergence close to zero  
(KL high, KL low = 0)
- reconstruction loss decrease

how to prevent:

- increases the weight of the KL term during early training.  
a small  $\beta$  allows the model to focus on reconstruction without forcing the posterior to match the prior too strictly

• Interpretation of what each latent dimension learned to control

-  $z_{\text{high}}$  dimensions control

global style

-  $z_{\text{low}}$  dimensions control

local variations (kick, snare, ...)

• Quality assessment:

Do generated patterns sound musical?

Yes, I use the function

`drum_pattern_validity()` provided

in homework to check if the

patterns are musically valid,

And the results are 1.

## \* Comparison of different annealing strategies

### • KL annealing

start with a small KL weight

so the model focuses on reconstruction

later increase  $\beta$  to 1 as training progresses,

forcing the model to

structure the latent space

### • Free bits

impose a minimum KL threshold

per dimension, KL below this

threshold is clamped to ensure

each dimension carries information

instead of being ignored

## - Temperature annealing

High temperature  $\rightarrow$  softer,  
more random output

low temperature  $\rightarrow$  sharper,  
more deterministic  
outputs

decrease temperature during training  
for sharper final generation