3 week 12

16 MON indices: BxT

17 TUE encode return quantized, indices
BxTxD BxT

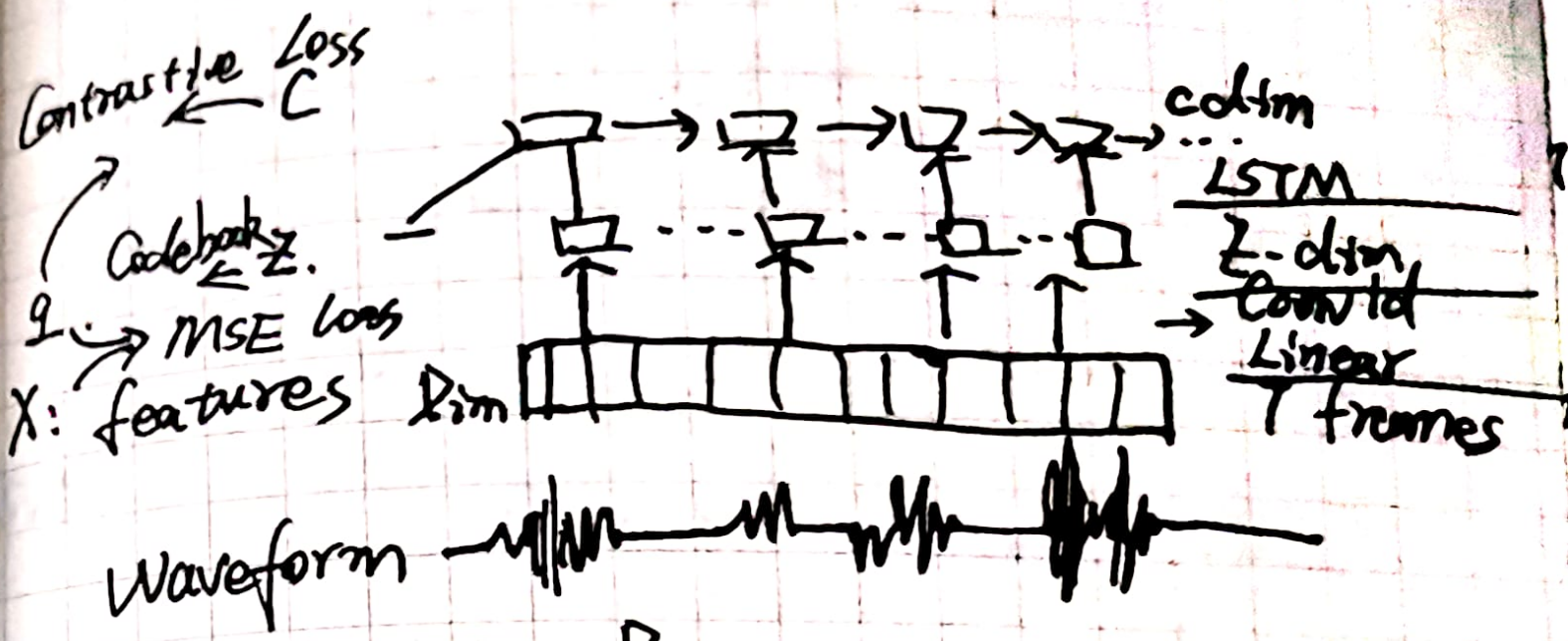
18 WED 或将indices → F.one_hot Ln类
后再embedding.

19 THU

20

21 AT

VQ-CPC



→ n × Z dim [n × D]

Codebook 与 Z 的 distance 由 torch.addmm 实现

out = β × mat + α(mat1 × mat2)

Sampling e ~ Uniform [-1/编, 1/编] [n × D]

mat = $\|Z\|_2^2 + \|e\|_2^2$ β = 1 α = -2

mat1 = Z mat2 = e^T 欧式距离

Z [B × T × D] e: [n × D]

distance: [B × T × n]

codebook 中

indices = argmin(distance, -1)

quantized: [B × T × D] → [B × T × D]

6 MON n_negatives：负样本数为 ng

7 TUE batch-index： u × ng，从 0、u 随机采样

reshape → [1 × u × ng × 1]

8 WED seq-index： s × u × ng × T-k 从 0, length 随机采样

9 THU f ← cat(z shift, z-negative) × Linear

10 FRI (C)

因 [B × T-k × c-dim] 果本身与 negative

11 SAT 同类，此时用交叉熵损失，与

label [s×u, ng+1, :] 全 0 计算

12 得 CPC Loss

CPC Loss Input: n_prediction_steps (K)

z n_negatives, z_dim

predictor. c_dim

steps (K) × Linear[c_dim, z_dim]. n_speakers_per_batch.

n_utterances_per_spk

forward. Input: z [B×T×z_dim]

speaker c [B×T×c_dim]

z → Reshape [S×u×T×z_dim]

u: utterance.

c 裁掉K帧 → [B×T-K×c_dim]

↓Predictor

c: Length

z: Prediction Steps K

for i=1…k

z_shift

[S, u, T移z dim