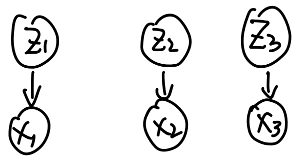
GMM

Dataset $= \{x^{(1)}, \dots, x^{(n)}\}$

Each $x^{(i)}$ comes from a latent cluster z

Each z_i drawn independently

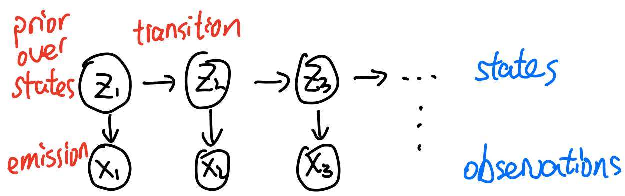


HMM

Data sequence: $x_1, \dots, x_T$ T = total timesteps

Each observation x_t has latent state z_t

Each z_t depends on previous state z_{t-1}



Casino dice game

Fair → Fair → F → Loaded → L → L → L → L → F → F

2 5 3 6 6 4 6 6 1 2

state : fair dice or loaded dice

observations : dice rolls

Genomics

Coding C C C N N N C C

A T G C T A G C A

state: coding vs. non-coding

observations : DNA base pairs

Speech recognition

rec	og	nize	speech
wreck	a	nice	beach

states : words / symbols

observation : audio

HMM (formal): probabilistic model of a sequence of observations $x_1, \dots, x_T$

① Sample z_1 from prior distribution $\pi_{1:k}$ (k possible states)

$$p(z_1 = c) = \pi_c$$

② For each $t = 2, \dots, T$

sample z_t from $p(z_t \mid z_{t-1})$

represent as $k \times k$ matrix A where $a_{ij} = p(z_t = j \mid z_{t-1} = i)$

$$\begin{array}{c|cccc} i= \backslash j= & 1 & 2 & 3 & 4 \\ \hline 1 & & & & \\ 2 & .1 & .3 & .4 & .2 \\ 3 & & & & \\ 4 & & & & \end{array}$$

Markov assumption: z_t only depends on z_{t-1}

Also assume same transition probabilities (same A) at each timestep

③ For each $t = 1, \dots, T$

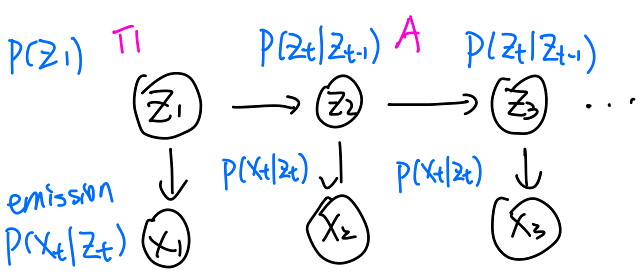
sample x_t from $p(x_t \mid z_t)$

Assume that x_t only depends on z_t

Assume emission dist is same $\forall t$

Genomics: $p(x_t \mid z_t)$ is dist over $\{A, T, C, G\}$ for each value of z_t

Speech: $p(x_t \mid z_t)$ is Gaussian in audio feature space.



Inference in HMM

inferring values of latent variables ($z_1 \cdots z_T$) given observations ($x_1, \dots, x_T$)

assume we know:

1. π is $p(z_1)$
2. A is $p(z_t | z_{t-1})$
3. $p(x_t | z_t)$

Given $x_1, \dots, x_T$, try to infer

(A) the most likely sequence of $z_1, \dots, z_T$

$$\underset{z_{1:T}}{\arg\max}\ p(z_{1:T} | x_{1:T})$$

e.g. given audio, most likely words

(B) at a particular time t, what is the dist of z_t?

$$p(z_t | x_{1:T})$$

e.g. in casino, how likely is player cheating at time 9

A: Viterbi algorithm

$$\underset{z_{1:T}}{\arg\max}\ p(z_{1:T} | x_{1:T}) = \underset{z_{1:T}}{\arg\max}\ p(z_{1:T}, x_{1:T})$$

differ by a factor $P(x_{1:T})$

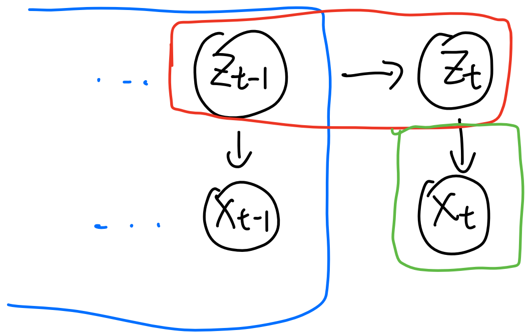
Strategy: dynamic programming

$$m_{tj} = \max_{\substack{z_{1:t} \\ \text{where} \\ z_t = j}} p(z_{1:t}, x_{1:t})$$

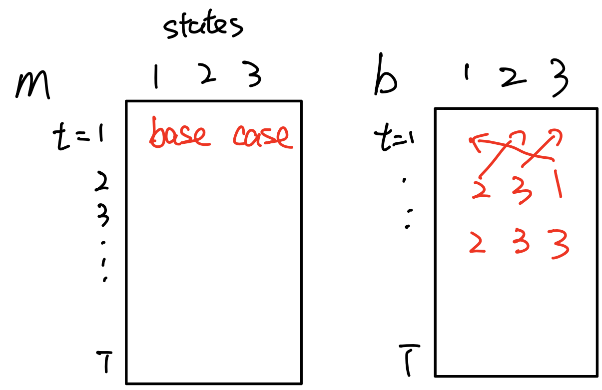
prob of best sequence of states upto time t where we end at state j

$$m_{tj} = \max_{i=1\cdots k} p(z_t = j \mid z_{t-1} = i)\, m_{t-1,i}\, p(x_t \mid z_t = j)$$

$$b_{tj} = \operatorname*{argmax}_{i=1\cdots k} p(z_t = j \mid z_{t-1} = i)\, m_{t-1,i}\, p(x_t \mid z_t = j)$$



m_{t-1} looks here



Base case $t=1$: $m_{1j} = p(z_1 = j)\, p(x_1 \mid z_1 = j) \quad \forall j = 1, \cdots, k$

For $t = 2, \cdots, T$

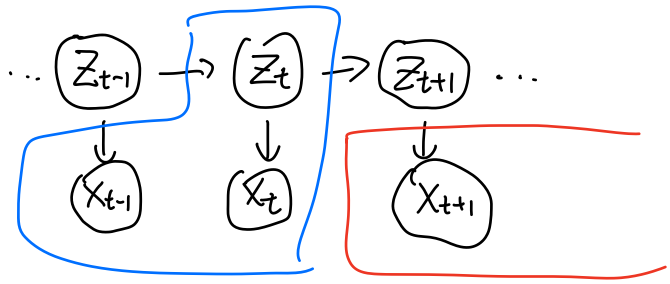
compute $\begin{cases} m_{tj} \text{ given } m_{t-1,j} \quad \forall j = 1, \cdots, k \\ b_{tj} \end{cases}$

At the end:

best ending state $z_T = \operatorname*{argmax}_{j=1\cdots k} m_{Tj}$

extract best path by following b_{tj}'s

(B) inference on Z_t



to infer $P(Z_t | X_{1:T})$

we need to reason

① Z_t's effect on X_t

② How past influences Z_t

③ How does Z_t affect future

future states

$$P(Z_t | X_{1:T}) : \frac{P(Z_t, X_{1:T})}{P(X_{1:T})} = \frac{P(X_{1:t}, Z_t)\, P(X_{t+1:T} | Z_t)}{P(X_{1:T})}$$

compute recursively

normalizing constant

compute numerator for all choices of Z_t

then normalize