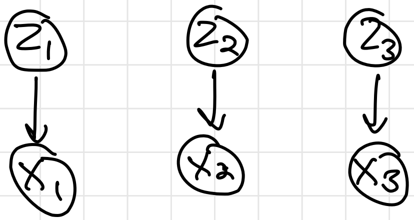
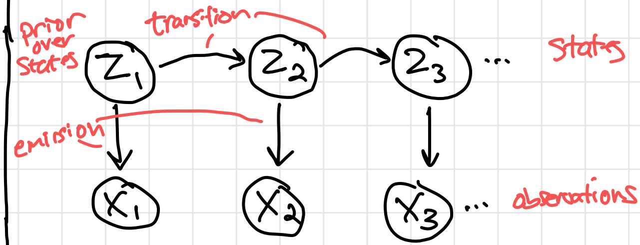
3/28/2023 Hidden Markov Model (HMM)

T = total timesteps

GMM	HMM
Dataset $= \{x^{(1)}, \dots, x^{(n)}\}$	Data sequence: $x_1, \dots, x_T$
Each $x^{(i)}$ comes from a latent cluster z_i:	Each observation x_t has latent state z_t
Each z_i drawn independently	Each z_t depends on previous state z_{t-1}





Casino dice game

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

2 5 3 $\underbrace{6\ 6\ 4\ 6\ 6}$ 1 2

state: Fair dice or loaded dice

observations: dice rolls

Genomics

Coding C C C N N N N C C

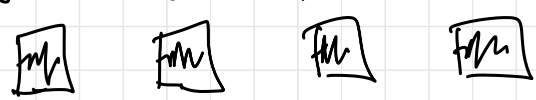
A T G G C T A G C A

State: coding vs non-coding

observations: DNA base pairs

Speech Recognition

ree	-og	-nize	speech
wreck	a	nice	beach



States: words/syllables

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 $A_{ij} = p(z_t = j \mid z_{t-1} = i)$

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

Markov Assumption:

z_t only depends on z_{t-1} not on other history

Also assume same transition probabilities (ie same A) at each time

③ For each $t = 1, \dots, T$

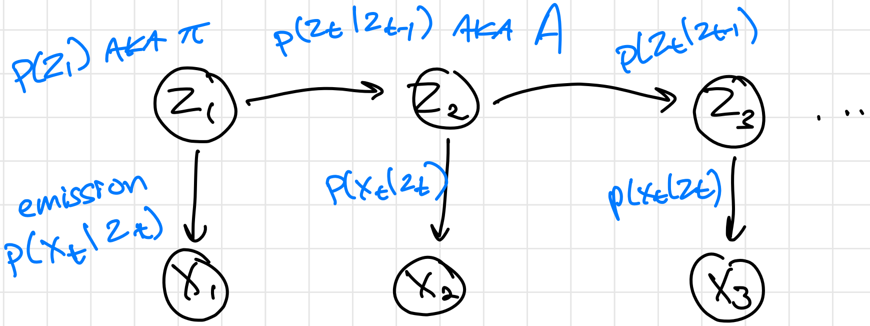
Sample x_t from $p(x_t \mid z_t)$

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

Assume that x_t only depends on z_t (not past states or observations)

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

Assume emission distribution is same for all times t



HW3: Out tonight

Inference in HMMs

Inferring values of latent variables $z_1, \dots z_T$ given observations $x_1, \dots, x_T$

Assuming we know:

1. π ie $P(z_1)$
2. A ie $P(z_t \mid z_{t-1})$
3. $P(x_t \mid z_t)$

Next class: learn these

Given observations $x_1, \dots, x_T$ we could try to infer:

(A) Most likely sequence of $z_1 \dots z_T$

$$\operatorname*{argmax}_{z_{1:T}} P(z_{1:T} \mid x_{1:T})$$

e.g given audio, find most likely sequence of words

(B) At a particular time t, what is distribution of z_t?

$$p(z_t \mid x_{1:T})$$

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

A: Viterbi Algorithm

$$\operatorname*{argmax}_{z_{1:T}} P(z_{1:T} \mid x_{1:T}) = \operatorname*{argmax}_{z_{1:T}} P(z_{1:T}, x_{1:T})$$

Differ by factor of $P(x_{1:T})$

Strategy: Dynamic Programming

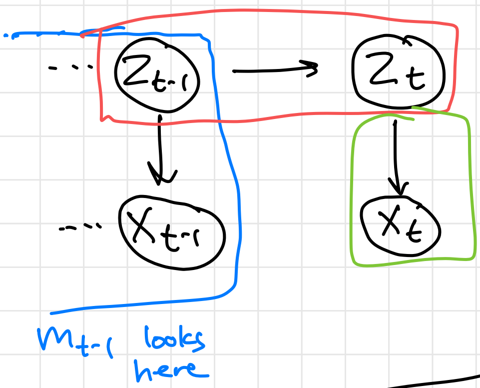
$$m_{tj} = \max_{\substack{z_{1:t} \\ \text{where} \\ z_t = j}} P(z_{1:t}, x_{1:t})$$

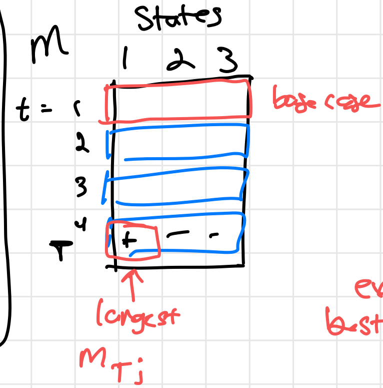
Probability of best sequence of states up to time t where we end at state j

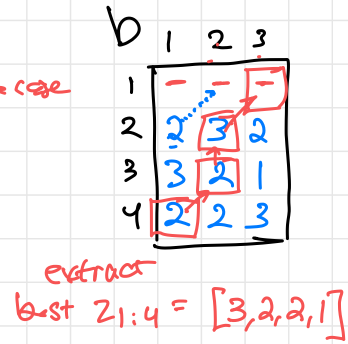
$$m_{tj} = \max_{i=1, \dots, K} p(z_t = j \mid z_{t-1} = i)\, m_{t-1,i}\, p(x_t \mid z_t = j)$$

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

"backpointers"







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

For $t = 2, \dots, T$

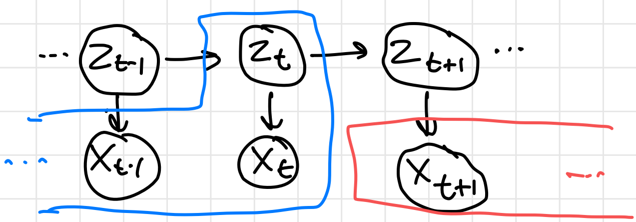
Compute m_{tj} given $m_{t-1,j}\; \forall j = 1, \dots, k$
and b_{tj}

At the end:

Best ending state $z_T = \underset{j=1 \dots k}{\operatorname{argmax}} \; m_{Tj}$

extract best path by following b_{tj}'s

(B) Inference on z_t



To infer $p(z_t \mid x_{1:T})$
we need to reason about:

1. z_t's effect on x_t
2. How past influences z_t
3. How z_t affect future

$$p(z_t \mid x_{1:T}) = \frac{p(z_t, x_{1:T})}{\boxed{p(x_{1:T})}} = \frac{\boxed{p(x_{1:t}, z_t)}\,\boxed{p(x_{t+1:T} \mid z_t)}}{p(x_{1:T})}$$

Normalizing constant.
Compute numerator for all choices of z_t
then normalize

Compute recursively