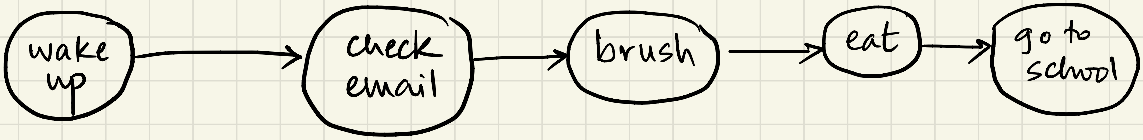
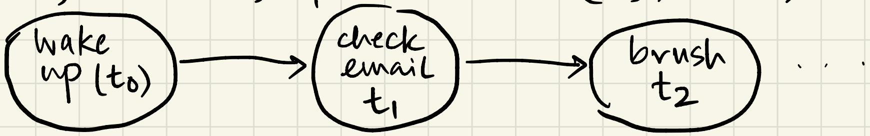
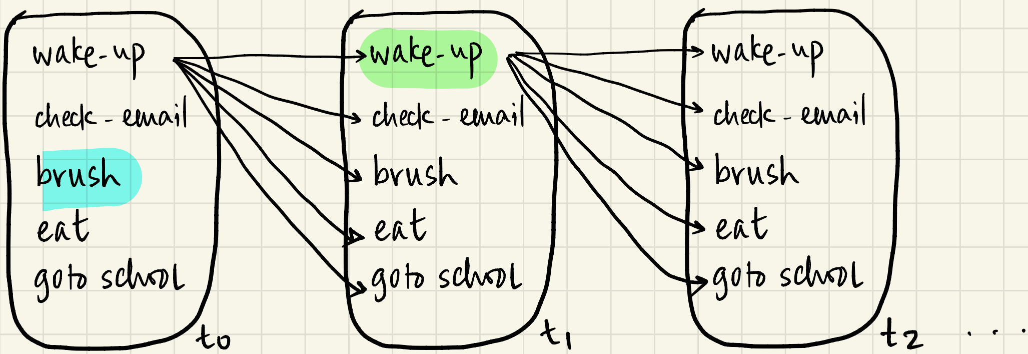
Probabilistic Graphical Models

① Modeling sequence of events with graphs (nodes, edges)

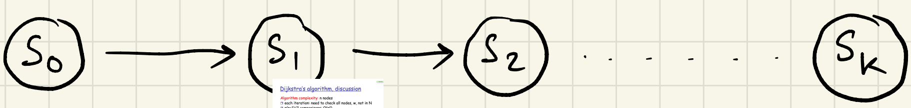


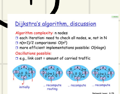
- Edges convey progression (e.g., time)
- 

② Probabistic graphical models → each node is now a random variable (RV)

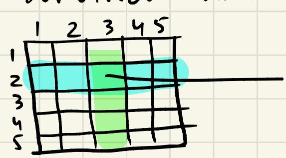


- Let's write the modeled state at time t_K as S_K





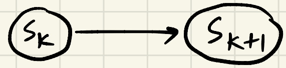
- Transition [illegible] one state to another has probabilities. Let's model as matrix



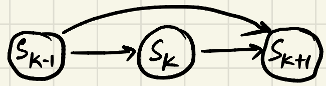
P_{23} = brush to wake

④ Modeling short / long term dependencies

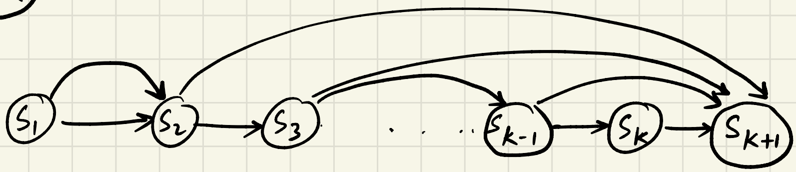
- should s_{k+1} only depend on s_k?



Or perhaps s_{k+1} also may depend on s_{k-1}



Or even all $s_{1:k}$ since the whole build up matters



- Answer depends on the application

A. s_{k+1} = Final SAT score
s_k = k^{th} practice SAT score
s_{k-1} = $(k-1)^{th}$ practice SAT score ...
$\vdots$

B. s_{k+1} = Next location of a flying bee
s_k = Current location
s_{k-1} = Previous location
$\vdots$

④ Mathematically speaking : Conditional Independence

$$P(X,Y|Z) = P(X|Z)\,P(Y|Z)$$

$$\equiv P(X|Y,Z) = P(X|Y) \equiv P(s_{k+1}|s_k, s_{k-1}) = P(s_{k+1}|s_k)$$

- Please prove this if s_k is conditionally independent of s_{k-1}

- Physical meaning of conditional independence?

$$P(\text{height}, \text{Vocabulary}) \stackrel{?}{=} P(\text{height})\,P(\text{vocab})$$

$$P\left(h, v \,\middle|\, \begin{array}{c}\text{students}\\ \text{of this}\\ \text{class}\end{array}\right) = P\left(h \,\middle|\, \begin{array}{c}\text{students}\\ \text{of this}\\ \text{class}\end{array}\right) P\left(v \,\middle|\, \begin{array}{c}\text{students}\\ \text{of this}\\ \text{class}\end{array}\right)$$

- Any other example?

5 Markovian or Non-Markovian?

- When $P\left(S_{k+1} = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}\right)$ only a function of S_k, then called "Markovian"

i.e., the bee is memoryless, hence Markov process.

- When calculating $P\left(S_{k+1} = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}\right)$ depends on multiple prior states, then non-markovian.

i.e., $P(SAT_{k+1} \mid SAT_k, SAT_{k-1}, \ldots SAT_{k-r}) \neq P(SAT_{k+1} \mid SAT_k)$

- How do you know Markovian or not?
  - → Experience / intuition
  - → Data driven calculations (offline)

(6) Examples : Memoryless or not ?

(a) $P(\text{destination} \mid \text{now at Zurich airport}, \text{departure airport})$

(b) $P(\text{job interview} \mid CGPA_8, CGPA_7 \ldots CGPA_1)$

$\overset{?}{=} P(\text{job interview} \mid CGPA_8)$?

* (c) $P(\text{Ball location in the goal}_{t+1} \mid \text{Ball location}_t, \text{Ball loc.}_{t-1}, \ldots \text{Ball Loc.}_{t_0})$

$\overset{?}{=} P(\text{Ball loc. in goal}_{t+1} \mid \text{Ball loc.}_t)$

(d)

h	e	l	l	?
s_1	s_2	s_3	s_4	s_5

l	?
s_4	s_5

$P(s_5 \mid s_1\ s_2\ s_3\ s_4) \overset{?}{=} (s_5 \mid s_4)$

7 Next state prediction (Markov)

Need to estimate **Transition Probability Matrix** (TPM)

$$S_k \xrightarrow{P_{ij}} S_{k+1}$$

S_k = \ S_{k+1} =	beckman	eceb	siebel	cif	union
beckman					
eceb					
siebel					
cif					
union					

This TPM is often a fn of a **model**.

A rocket moving as a fn of fuel injection and air drag

A person typing in English / other language and the next character has some TPM as a function of that language.

- **Predicting** or **tracking**

Predicting: Predict S_{k+1} from $S_{1:k}$

Tracking: Predict $S_{1:k}$ from $S_{1:k}$

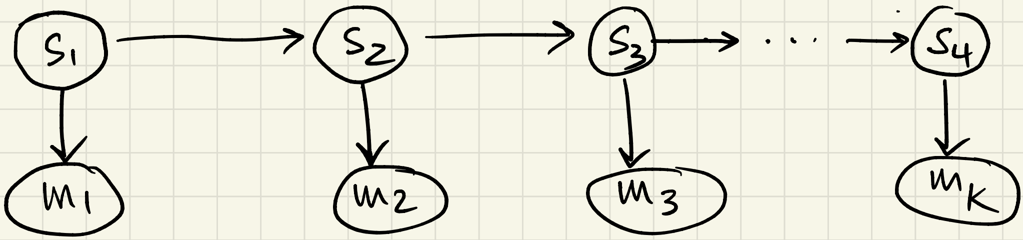
Can you track the rocket or predict the typed sentence?

⑧ Hidden States and Emission

- Tracking from the process alone may be difficult in some applications ... impossible in others

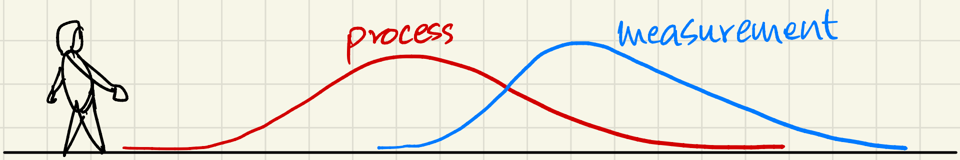
  ↳ Because no way to correct wrong $\hat{s}_k$

- Some measurements needed.



- Measurements (also called emissions) help in "resetting" or "recalibrating" error
- Measurements need not be accurate

  ↳ i.e., also a random variable

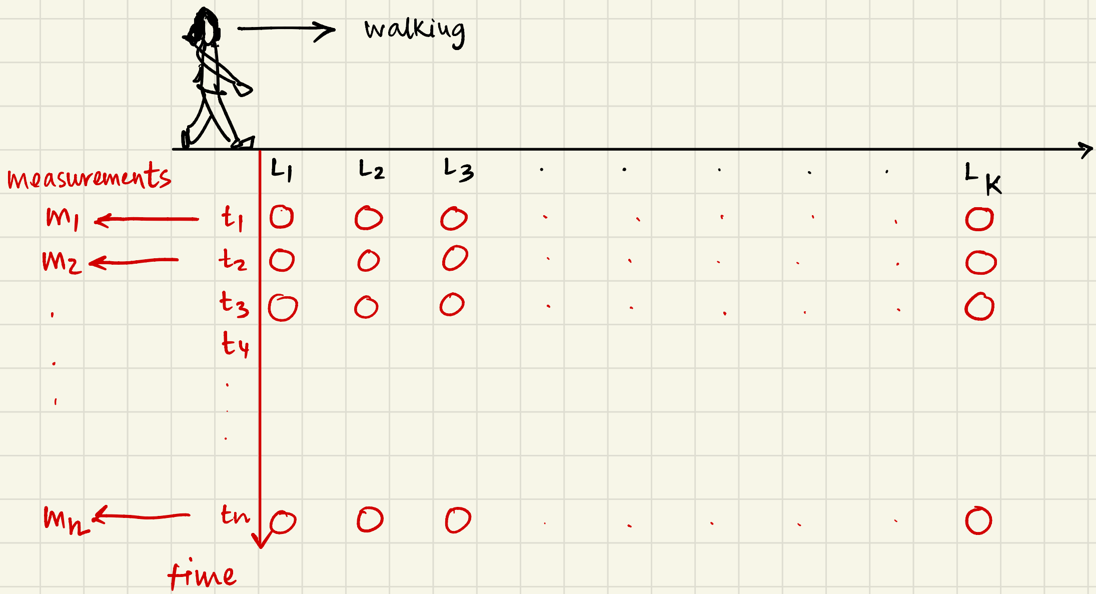


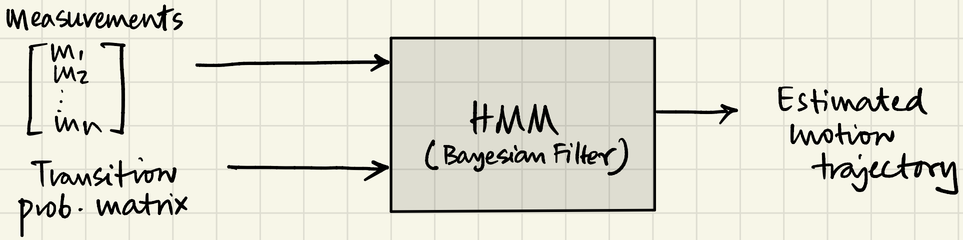
- Measurements need not be of s_k,

  ↳ i.e., can be $m_k = A s_k + \text{noise}$

Hence, s_k is hidden. Some emission or leakage gives us m_k. Now predict s_k.

HIDDEN MARKOV MODELS (HMM)





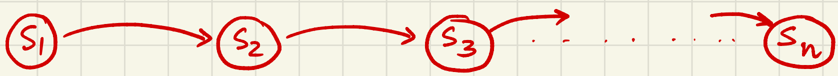
⑦ Formulating the state transition diagram :

Let S_k denote the state of the subject at time k

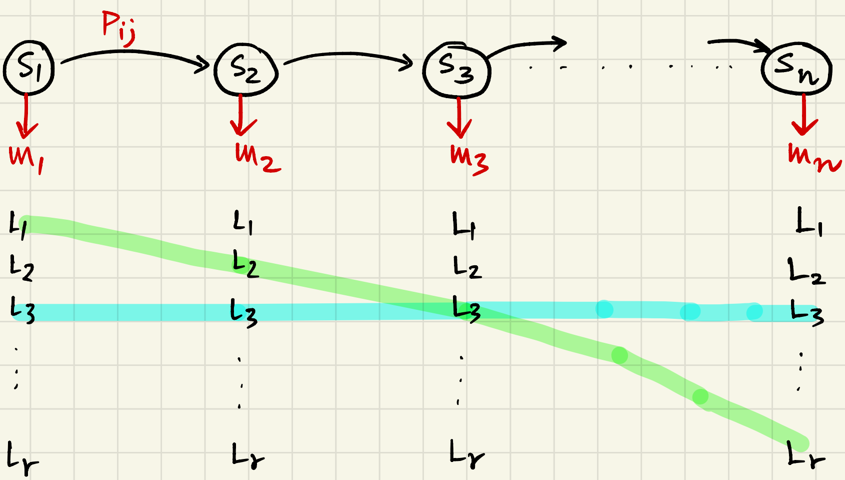
→ S_k is a random variable, i.e., $S_k = \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_r \end{bmatrix} \quad \forall k \in [1, n]$

→ S_k called the "state variable"

The human's walking motion is captured in



And the measurement and motion model is available for each state



⑧ Key Question : Where is/was the human at time t_k ?

$P(S_k \mid m_{1:k})$ or $P(S_k \mid m_{1:n})$

Is this : Posterior or likelihood ?

Do you have an intuitive feel for $P(s_k \mid m_{1:n})$?
If not, fall back on visualizing them as vectors

$$P\left(s_k = \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_r \end{bmatrix} \middle| \; m_1 = \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_r \end{bmatrix} \; m_2 = \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_r \end{bmatrix} \cdots m_n = \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_r \end{bmatrix} \right)$$

$$= \frac{P(s_k, m_{1:n})}{P(m_{1:n})} =$$

- From this joint distribution (in numerator) you want to know which value of $s_k = \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_r \end{bmatrix}$ has the max probability given the $m_{1:n}$ measurements you already have.

- The denominator is same for all $s_k = \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_r \end{bmatrix}$, so only numerator matters. But we don't know that joint distribution.

- Turn this posterior to likelihoods : $\dfrac{P(m_{1:n} \mid s_k)\, P(s_k)}{P(m_{1:n})}$

Likelihood ... and that is not hard because it's the sensor's measurement quality

Who cares! We only want to compare the numerator, so ignore denominator.

Hmmm! This depends on where I was last. So fn of $P(s_{k-1})$

- Let's do this mathematically now.

- Let's start with a basic result for $P(\underbrace{s_{1:n} \mid m_{1:n}}_{\text{trajectory given measurements}})$

$$P(s_{1:n} \mid m_{1:n}) = \frac{P(s_{1:n}, m_{1:n})}{P(m_{1:n})} \propto P(s_{1:n}, m_{1:n})$$

$$P(s_{1:n}, m_{1:n}) \overset{\text{chain rule}}{=} P(m_n \mid m_{1:n-1}, s_{1:n})\, P(m_{n-1} \mid m_{1:n-2}, s_{1:n}) \cdots$$

$$\cdots P(m_1 \mid s_{1:n})\, P(s_n \mid s_{1:n-1}) \cdots P(s_2 \mid s_1) P(s_1)$$

Markov

$$= P(m_n \mid s_n)\, P(m_{n-1} \mid s_{n-1}) \cdots P(m_1 \mid s_1)\, P(s_n \mid s_{1:n-1}) \cdots P(s_2 \mid s_1) P(s_1)$$

$$\boxed{P(s_{1:n}, m_{1:n}) = P(m_1 \mid s_1) P(s_1) \prod_{i=2}^{n} P(m_i \mid s_i)\, P(s_i \mid s_{i-1})}$$

- Now we want $P(s_k \mid m_{1:n})$

$$P(s_k \mid m_{1:n}) \propto P(s_k, m_{1:n}) = P(s_k, m_{1:k}, m_{k+1:n})$$

$$= P(m_{k+1:n} \mid s_k, m_{1:k})\, P(s_k, m_{1:k})$$

$$= P(m_{k+1:n} \mid s_k)\, P(s_k \mid m_{1:k})\, P(m_{1:k})$$

$$= P(s_k \mid m_{1:k}) \cdot P(m_{k+1:n} \mid s_k)$$

Forward (online) → Probability that murder suspect is at s_k = green st. given $k=4$ recent surveillance camera measurements of main street → wright street → 6th street

Backward (offline) → Probability that 4th to 8th measurement are Neil St. → Kirby road → Lincoln drive → university avenue, given suspect's k^{th} time location s_k = green street.

Questions

Prereqs

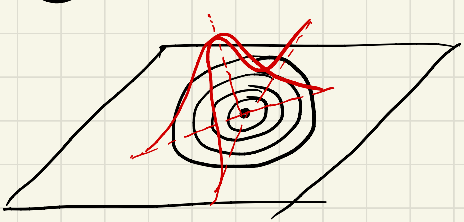
① Conditional prob. $P(X|Y,Z) = P(X|Y)$

$$P(X|Y,Z) = P(X,Z|Y)\cdot\frac{P(Y)}{P(Y,Z)}$$

$$= P(X|Y)\cancel{P(Z|Y)}\cdot\frac{\cancel{P(Y)}}{\cancel{P(Z|Y)}\cdot\cancel{P(Y)}}$$

$$= P(X|Y)$$

② Conditional distribution

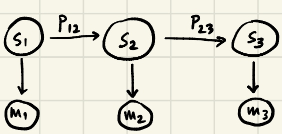


$$P\left(\text{measured gps location} \,\middle|\, \text{true location}\right) = \mathcal{N}\left(\text{true location}, \Sigma_1^2\right)$$

Not gaussian

Note $P\left(\text{true location} \,\middle|\, \text{measured location}\right) = ?$

because $P(\text{true}|\text{measured}) = \dfrac{P(m|t)\cdot P(t)}{P(m)}$



$m = As + n$

③ $Var(X+Y)$

$Var(KX)$

$\therefore Var(KX+Y)$

Also $Var(X+K)$

Need: P_{ij} $\quad P(m_i|s_i)$

Assume: $P(s_k|s_{k-1}\, s_{k-2}) = P(s_k|s_{k-1})$ & $P(m_k|s_{1:k}, m_{1:k-1}) = P(m_k|s_k)$