② 76

Hidden Markov Models (HMM)

Lets play a game. Alice & Bob are two friends. Based on Bob's mood (happy or grumpy), Alice got to predict the weather of today / tommorrow.

FACT: Bob is happy when it's sunny, he's grumpy when it's rainy. So, if Alice sees Bob happy, she infers that it must be sunny. OTOH, when Alice sees Bob grumpy, she infers that weather must be rainy.

Let's make it slightly more complicated. Say Bob is mostly happy when it's sunny & mostly grumpy when it's rainy. For ex:

Sunny → Bob's mood
- 0.8 (happy w/ prob of 80%)
- 0.2 (grumpy w/ prob of 20%)

Rainy → Bob's mood
- 0.6 (grumpy w/ prob of 60%)
- 0.4 (happy w/ prob of 40%)

Task back to Alice: Predict Bob's mood based on the new information.

3 days in a week of Bob's life: H, G, H, Predict: S, R, S (may be)

~~tat~~

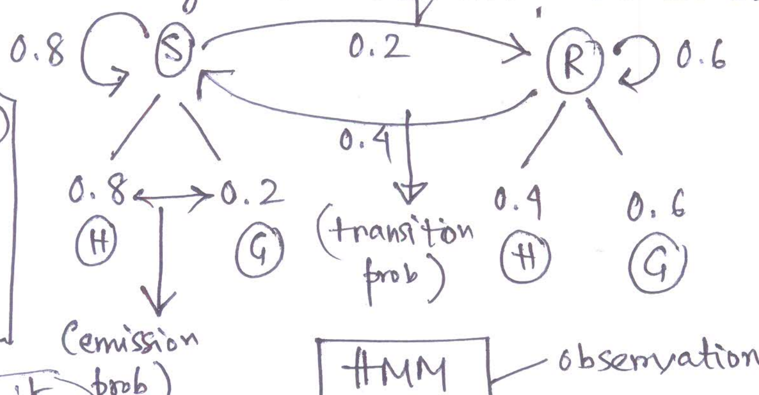
Predict Weather for the week

M	T	W	Th	Fri	Sat
H	G	H	G	H	G

predict → S → R → S → R → S → R (probably)

Is this a likely scenario? Weather doesn't behave like that!! How likely is this sequence? Let's add some additional information.

If today is S, prob (tom) = S is 80%, prob (tom) = R is 20%



If today is R, prob (tom = R) = 60% & prob (tom = S) = 40%.

these don't appear magically

HMM — observations ⇒ Alice sees Bob's mood, not weather

Transition prob: Prob of going from one state to another

Emission prob: Prob (observations made from the hidden states)

Questions

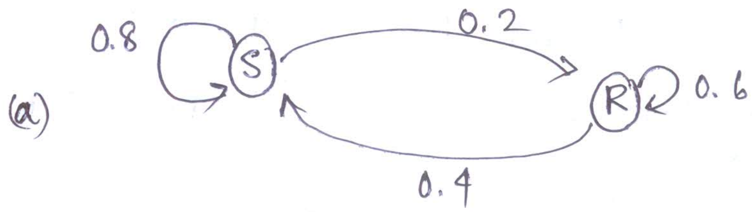
1. How did we find these probabilities?
2. What's the prob that a random day is S or R?
3. If Bob's happy today, what's the prob (Sunny or Rainy)?
4. If for three days Bob's H, G, H, what was the weather?

1:

G	H	H	H	G	G	H	G	H	H	H	G	H	H	H
S	S	S	S	R	R	R	S	S	S	S	R	R	S	S

← Fig. 1 historical data

How many times S is followed by S = 8 ⇒ 0.8
S is followed by R = 2 ⇒ 0.2
R is followed by S = 2 ⇒ 0.4
R is followed by R = 3 ⇒ 0.6



→ transition prob

S → H → 8 ⇒ 0.8
S → G → 2 ⇒ 0.2

R → H → 2 ⇒ 0.4
R → G → 3 ⇒ 0.6

Emission prob

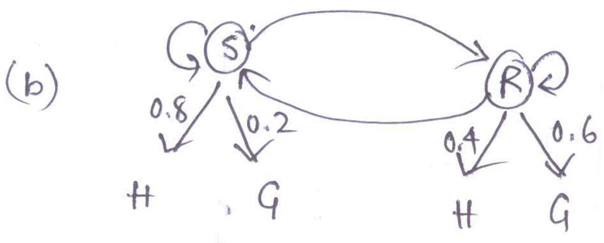


Fig 2(a) & (b)

2. What's the probability that a random day is (S) or (R)?

⇒ Say, Bob didn't give any information about his mood. How would Alice figure out? How likely is it that the weather is (S) or (R)?

Refer to fig 1 $\left[\begin{array}{l} (S) \rightarrow 10 \rightarrow \\ (R) \rightarrow 5 \rightarrow \end{array}\right\} \begin{array}{l} 2/3 \\ 1/3 \end{array}$

From Fig. 2a), we obtain

state equations $\begin{cases} S = 0.8S + 0.4R \\ R = 0.6R + 0.2S \end{cases}$ same eqn $+$ $\underbrace{S+R=1}_{\text{everyday is either (S)/(R)}}$

$\Downarrow$

$0.2S = 0.4R$

$\Rightarrow S = 2R \quad \Rightarrow \quad 2R + R = 1 \Rightarrow R = 1/3, \; S = 2/3$

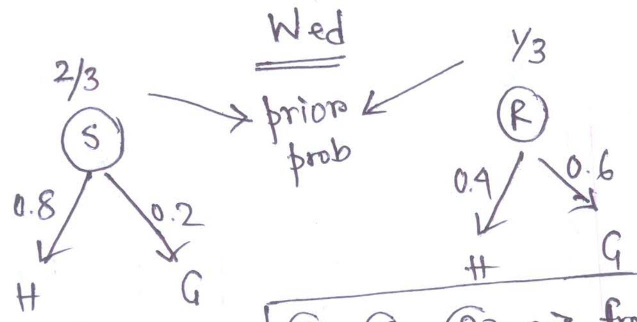
State eqns match observations!!

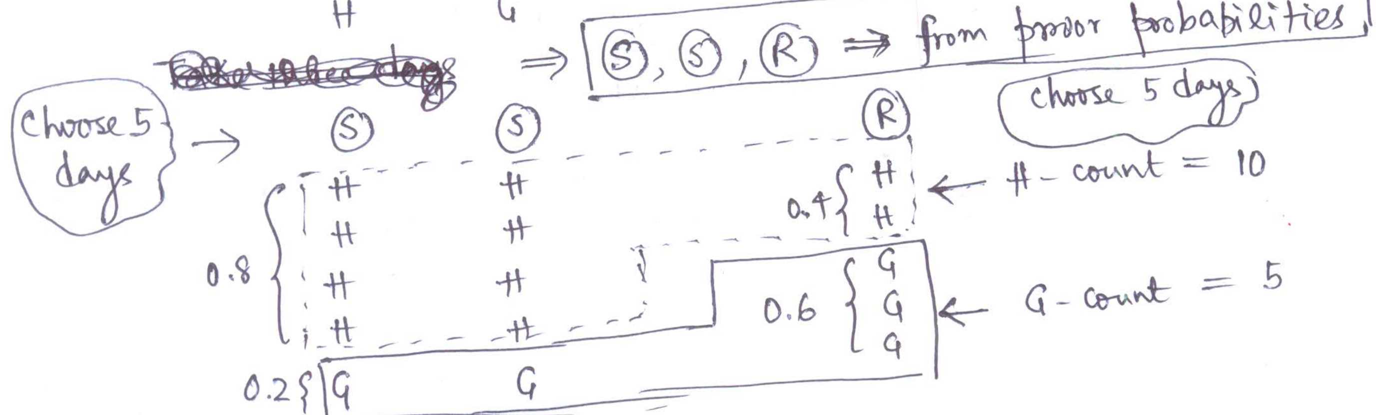
✓ It's 2/3 likely to be (S), 1/3 likely to be (R)

3. If Bob is (H) today, what's the prob ((S)/(R)) = ?

⇒ Bob gives information about his mood (Compare w/ Case 2)

Revisit Bayes Rule: If a day is (S) OR (R), that changes Bob's mood. But all of a sudden, we're given information about Bob's mood & we have to infer the weather ⇒ inversion ⇒ switching the events ⇒ Bayes rule.





Among the 10 happy days;

Next, $P(S \mid H) = 8/10$ ← 8 of these 10 days are S, other 2 are R

what is the prob that it's S ?

$P(R/H) = 2/10$ } posterior prob W/

Among the 5 Grumpy days:

posterior prob → $P(S/G) = 2/5$, $P(R/G) = 3/5$ } 2 of these 5 G days are S, other 3 are R

4. If for three days, Bob's H, G, H, what was the weather?

⇒ Let's take 2 days, H – G, what's the weather?

Wed	Thu
H	G
S R	S R

We have 4 scenarios

Consider this ⇒

Wed	→	Thu
S	→	S
S	→	R
R	→	S
R	→	R

} what's the prob that given these weather, Bob's H OR G and take the one w/ the highest value ⇒ MLE

⇓

	Wed		Thu
	H		G
	↑ 0.8		0.6 ↑
prior → 2/3 ←	S	$\xrightarrow{0.2}$	R

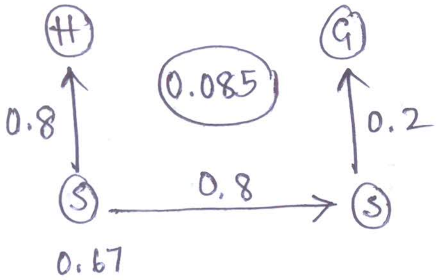
prob (Bob H / given S) = 0.8

P (Wed = S) = ?

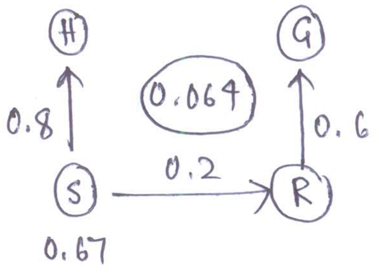
prob (Bob = G / Thu = R) = 0.6

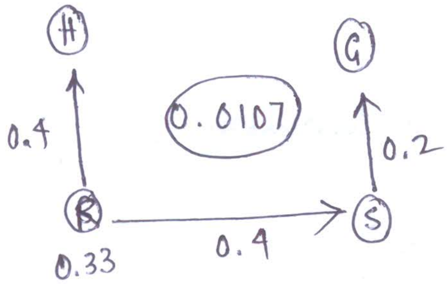
prob (Thu = R / wed = S) = 0.2

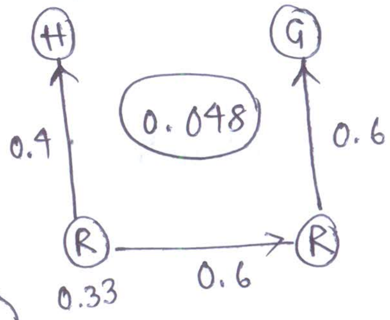
Prob (H – G) = product of all these (by def of conditional prob) = 0.0634

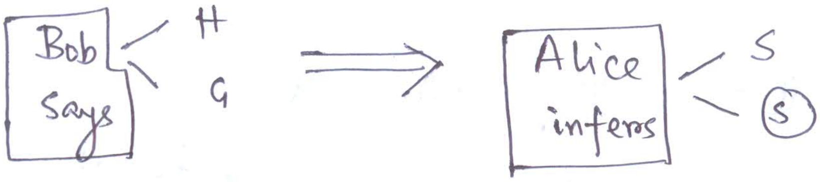


Highest
=
MLE
$\Downarrow$
Prob (Wed-Thu/H-G)
= S-S
is the
most
likely
$\Downarrow$
Weather
on Wed-Thu
= S-S

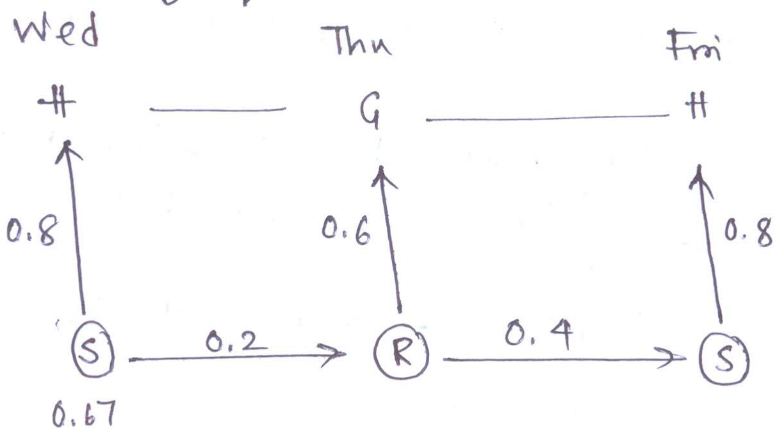








Now, consider a 3-day sequence Wed-Thu-Fri (8 scenario's)



Prob = 0.02048, check for all 8 Cases $\Rightarrow$ S-S-S

Homework $\longrightarrow$ final solⁿ

4 day scenario — 16 scenarios $= 2^4$

" n " — 2^n $\longrightarrow$ Let's do a formal framework.

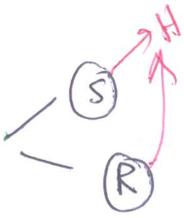
HMM

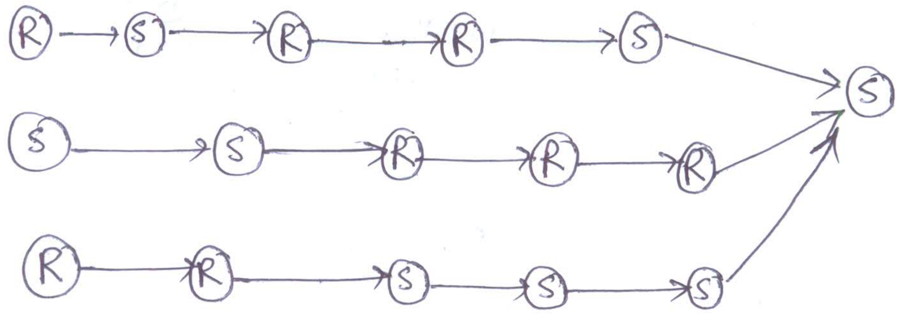
M	T	W	Th	Fr	Sa
H	H	G	G	G	H

S → R → S → R → S → R

the absurd case, remember?

Viterbi Algorithm: Tracing paths to a conclusion





Note: We only care about a sequence w/ max probability ⇒ we find out the best path to S & trace back the prob to H vs best path to R & trace back the prob to H

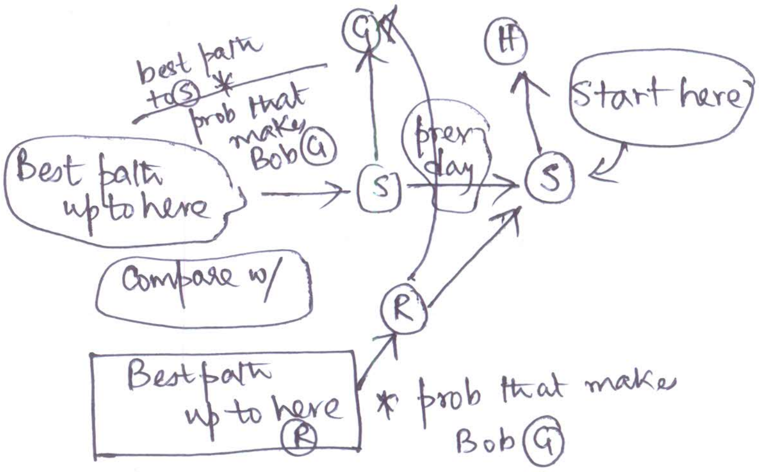
⇓

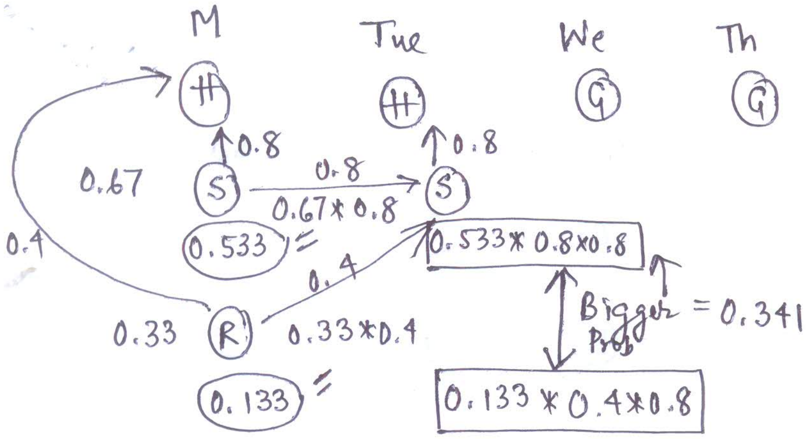
Prob (S day make Bob H) & Prob (R day makes Bob H)

& ~~take~~ select the one w/ max. prob.

O/P of Viterbi Algorithm ⟶ trace the best path to S & R

Tracing the path:





Tue → S / M → S ✓ selected

Tue → S / M → R & that makes Bob H

~~Prob~~

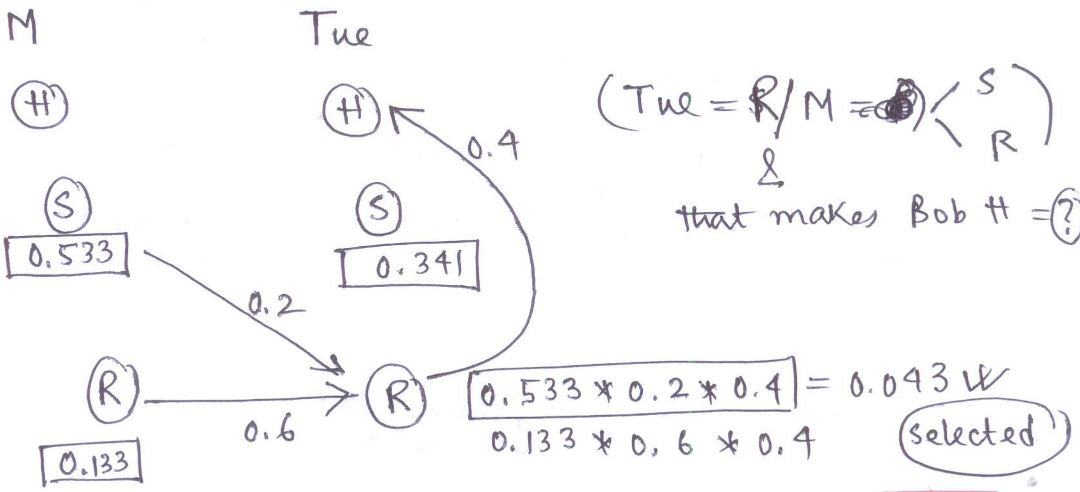
Seq 1 : A M → S → Tue → S   Prob (Bob is H / Tue = S / M = S) ✓ = 0.341

M → R

S – S & R – S } M – Tue

Seq 1 : B

S → R, R ← R } M – Tue

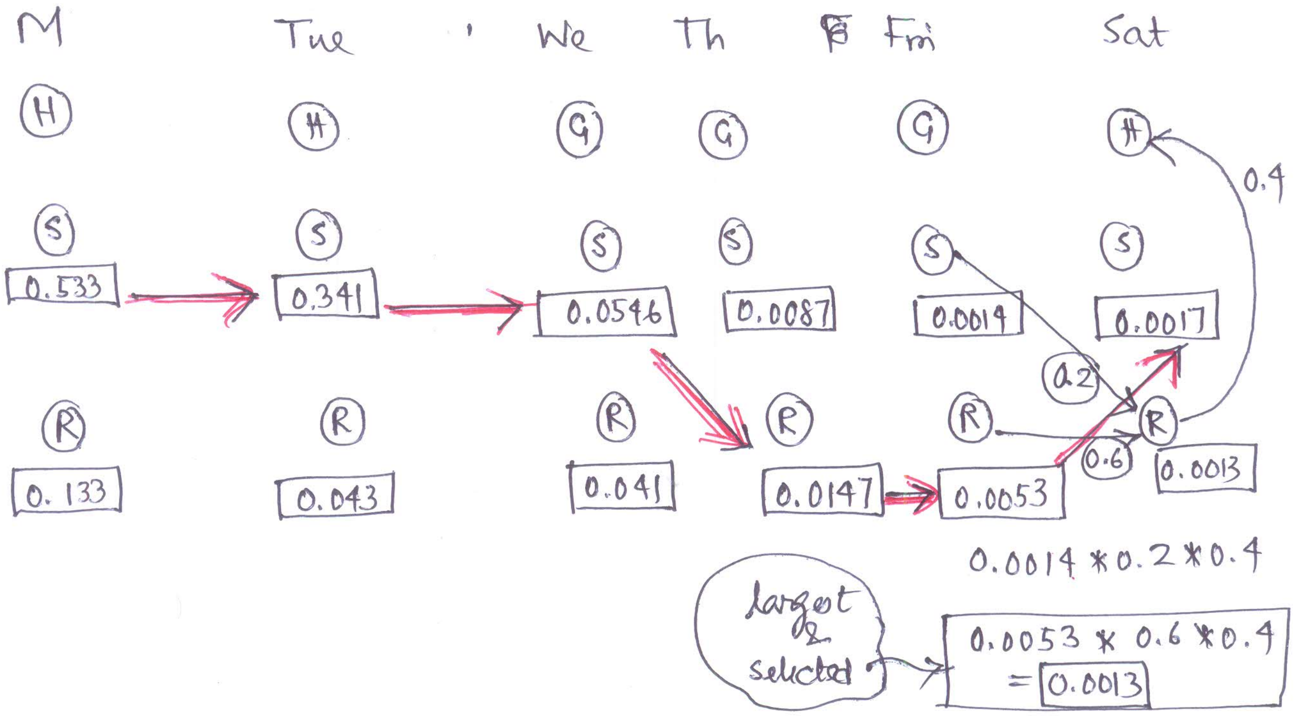


(Tue = R / M = S, R) & that makes Bob H = ?

Therefore Prob (Tue = R / M = S & that makes Bob H) = 0.043

∴ P(Bob = H / Tue – S, M – S) > P(Bob = H / Tue – R, M – S)

The algorithm goes this way... always picking the larger # ...



Answer:

M	Tue	We	Th	Fr	Sa
H	H	G	G	G	H
S →	S →	S →	R →	R →	S

So, if Bob's H → H → G → G → G → H, what's the weather?

S → S → S → R → R → S (most likely)

Note: (i) Difference b/w Markov chain & HMM is the emission probabilities ⇒ the correlation b/w observations & predicting hidden states !!

(ii) Markov chain possesses the property of independence of conditional probabilities ⇒ HMM exploits this property

= "Memoryless"