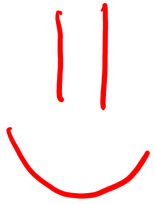
Warm-up as You Log In



Answer any query from the joint distribution

P(Weather)?

P(Weather | winter)?

P(Weather | winter, hot)?

Season	Temp	Weather	P(S, T, W)
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Announcements

Midterm 2

- Mon, 11/9, during lecture
- See Piazza for details

Introduction to Machine Learning

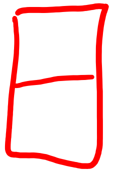
Hidden Markov Models

Instructor: Pat Virtue

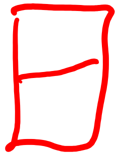
Warm-up as You Log In

Answer any query from the joint distribution

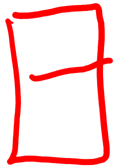
P(Weather)?



P(Weather | winter)?



P(Weather | winter, hot)?



Season	Temp	Weather	P(S, T, W)
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

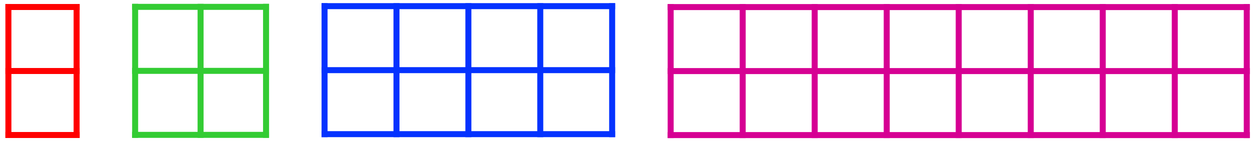
Bayesian Networks

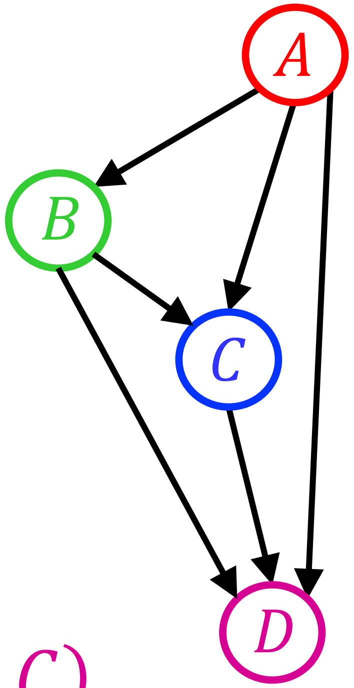
Bayes net

One node per random variable

Directed-Acyclic-Graph

One CPT per node: P(node | *Parents*(node))





$$P(A,B,C,D) = P(A)\,P(B|A)\,P(C|A,B)\,P(D|A,B,C)$$

Encode joint distributions as product of conditional distributions on each variable

$$P(X_1,\dots,X_N) = \prod_i P(X_i|\,Parents(X_i))$$

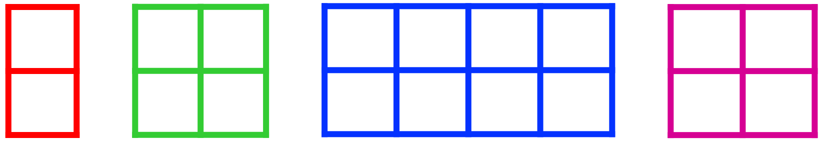
Bayesian Networks

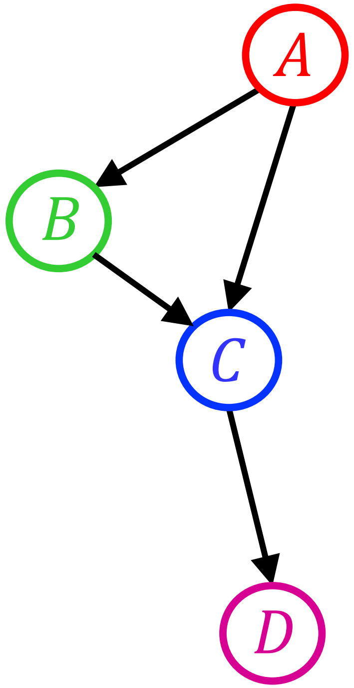
Bayes net

One node per random variable

Directed-Acyclic-Graph

One CPT per node: P(node | *Parents*(node))





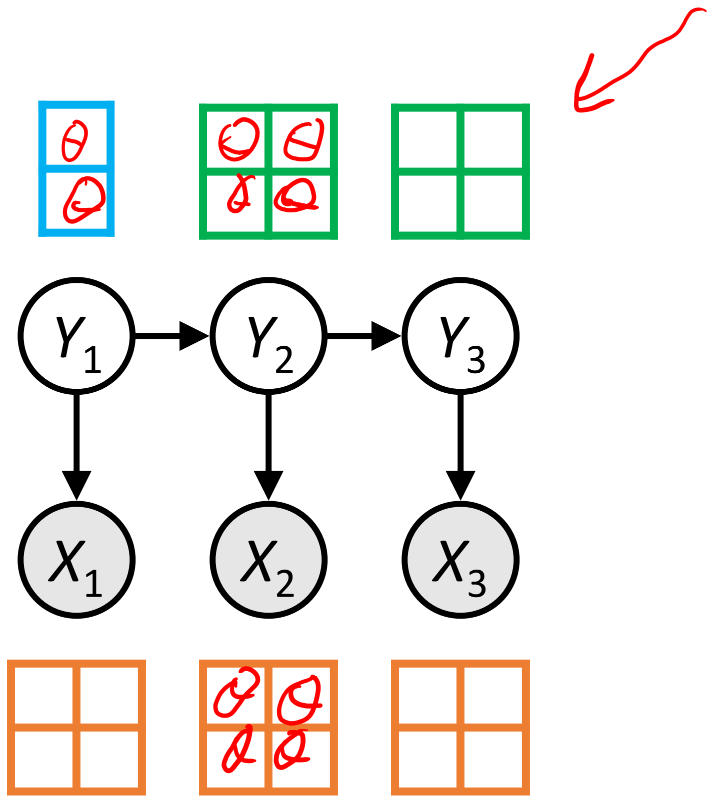
$$P(A,B,C,D) = P(A)\,P(B|A)\,P(C|A,B)\,P(D|C)$$

Encode joint distributions as product of conditional distributions on each variable

$$P(X_1,\ldots,X_N) = \prod_i P(X_i|\,Parents(X_i))$$

Outline

1. Probability primer
2. Generative stories and Bayes nets
   - Bayes nets definition
   - Naïve Bayes
   - Markov chains
   - Hidden Markov models
3. Learning HMM parameters
   - MLE for categorical distribution
4. Inference in Bayes Nets and HMMs

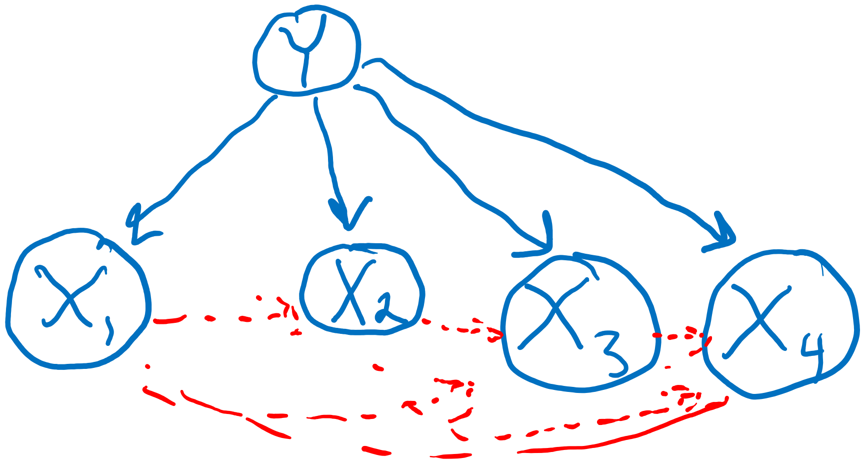


$P(Y_3 | x_1, x_2, x_3)$

Generative Stories and Bayes Nets

SPAM: Bag of words, naïve Bayes

- Generative story and Bayes net



$Y \sim Bern(\phi)$ → $P(Y=1)$

$X_{m,y=0} \sim Bern(\theta_{m,y=0})$ $P(X_m=1 \mid Y=0)$

$X_{m,y=1} \sim Bern(\theta_{m,y=1})$ $P(X_m=1 \mid Y=1)$

- Assumptions $X_i \perp\!\!\!\perp X_j \mid Y \quad \forall\, i \neq j$

- Joint distribution $P(Y)P(X_1 \mid Y)P(X_2 \mid Y)P(X_3 \mid Y)P(X_4 \mid Y)$

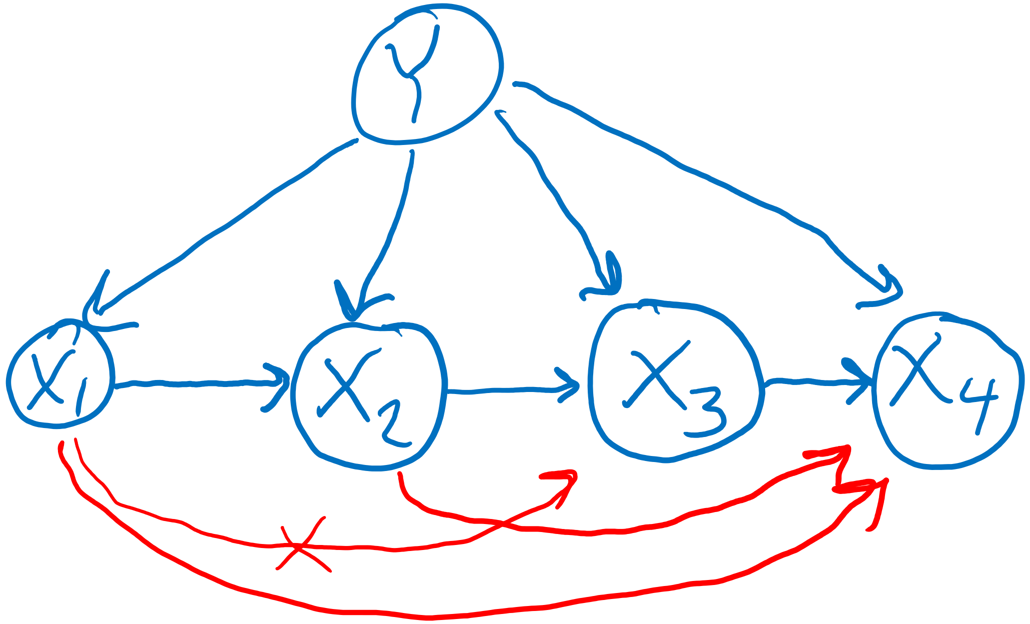
No assumptions:

$$P(Y, X_1, X_2, X_3, X_4) = P(Y)P(X_1 \mid Y)P(X_2 \mid Y, X_1)P(X_3 \mid Y, X_1, X_2)P(X_4 \mid Y, X_1, X_2, X_3)$$

Generative Stories and Bayes Nets

News article: Bigram

- Generative story and Bayes net



$$Y \sim Categorical(\boldsymbol{\phi})$$

$$X_{m,y=0} \sim Categorical(\boldsymbol{\phi}_{m,y=0})$$

$$X_{m,y=1} \sim Categorical(\boldsymbol{\phi}_{m,y=1})$$

- Joint distribution $P(Y)P(X_1 \mid Y)P(X_2 \mid Y, X_1)P(X_3 \mid Y, X_2)P(X_4 \mid Y, X_3)$

No assumptions:

$$P(Y, X_1, X_2, X_3, X_4) = P(Y)P(X_1 \mid Y)P(X_2 \mid Y, X_1)P(X_3 \mid Y, X_1, X_2)P(X_4 \mid Y, X_1, X_2, X_3)$$

Generative Stories and Bayes Nets

Weather

- Generative story and Bayes net

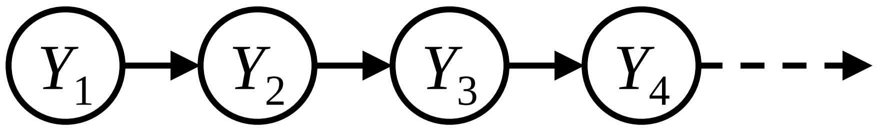
- Assumptions ?

- Joint distribution $P(W_1, W_2, W_3 \ldots) = P(W_1)P(W_2|W_1)P(W_3|W_2)\ldots$

Image: http://ai.berkeley.edu/

Markov Models

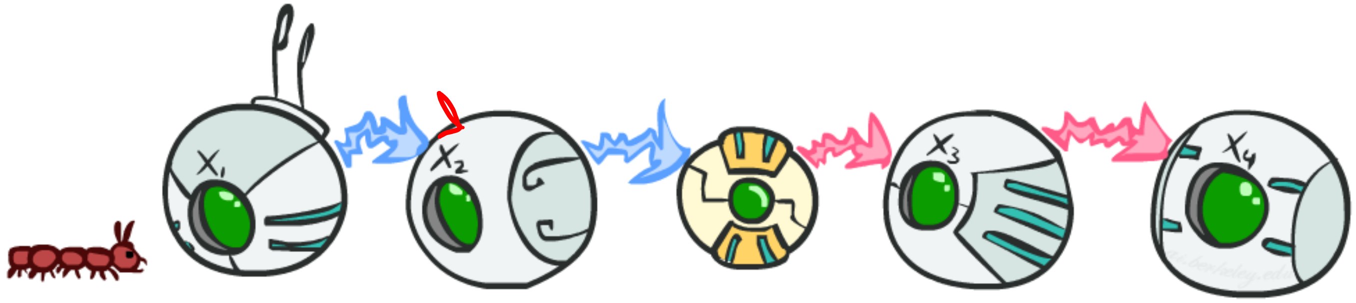
- Value of Y at a given time is called the state



$P(Y_1)$ $\qquad$ $P(Y_{t+1} \mid Y_t)$

- Parameters: called transition probabilities or dynamics, specify how the state evolves over time (also, initial state probabilities)
- Stationarity assumption: transition probabilities the same at all times

Markov Model Conditional Independence



Basic conditional independence:

- Past and future independent given the present
- Each time step only depends on the previous
- This is called the (first order) Markov property

$$Y_{t-1} \perp\!\!\!\perp Y_{t+1} \mid Y_t$$

$$Y_{t-2} \perp\!\!\!\perp Y_{t+1} \mid Y_t$$

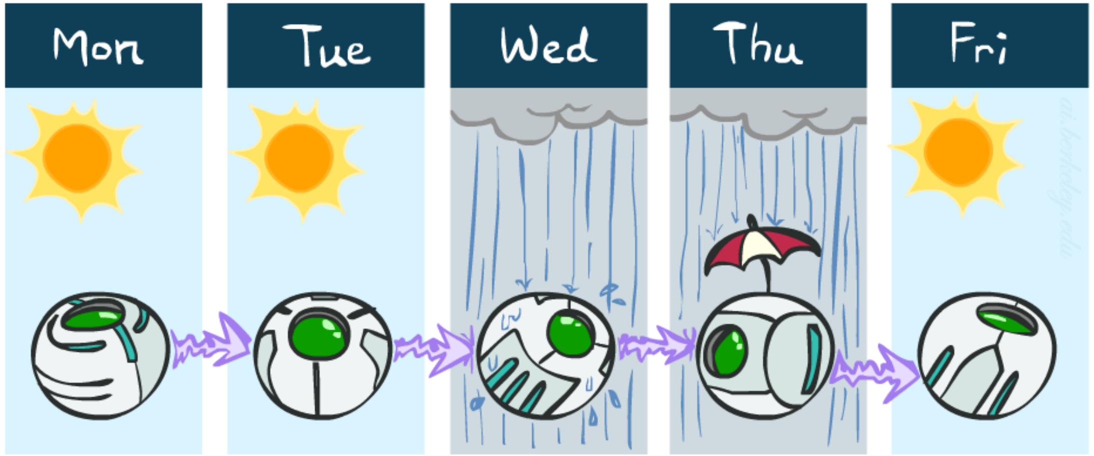
Image: http://ai.berkeley.edu/

Example: Markov Chain Weather

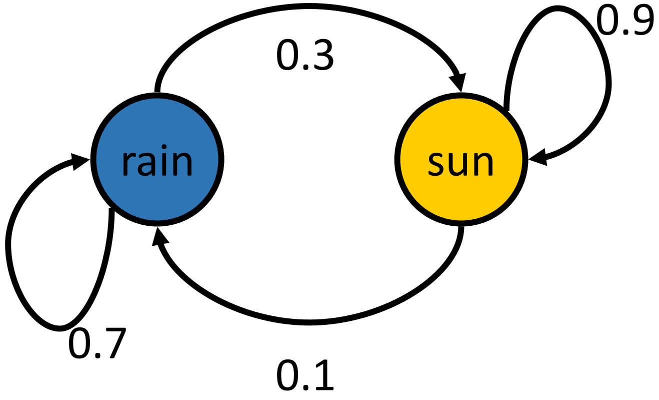
States: Y = {rain, sun}

- Initial distribution: 1.0 sun
- Conditional probability table (CPT) $P(Y_t \mid Y_{t-1})$:

Y_{t-1}	Y_t	$P(Y_t \mid Y_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7



Two other ways of representing the same CPT



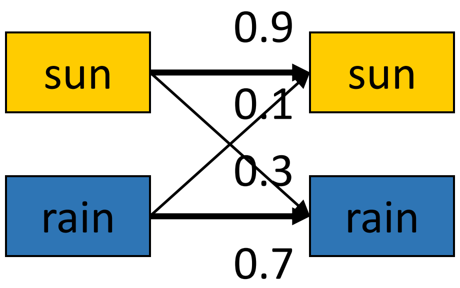
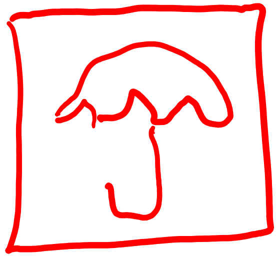


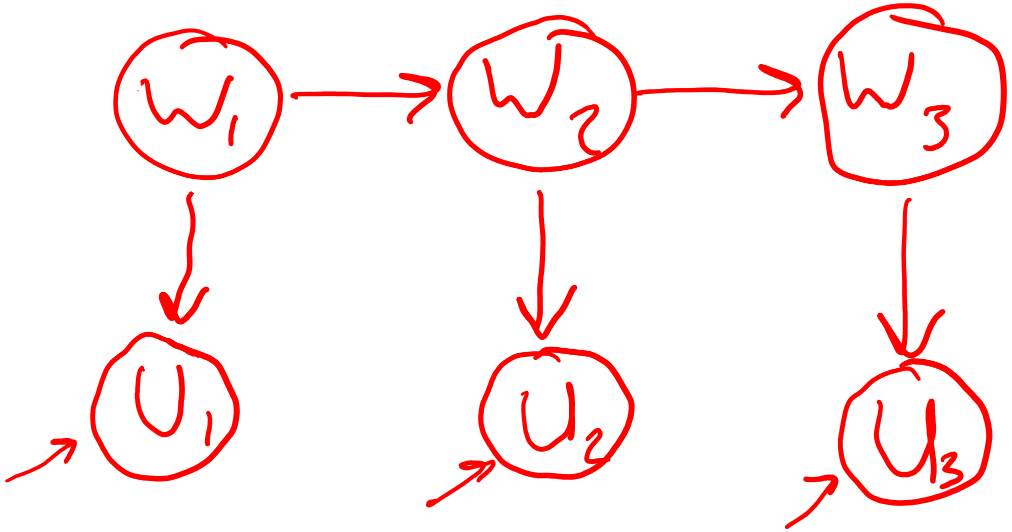
Image: http://ai.berkeley.edu/

Generative Stories and Bayes Nets

Weather, Umbrella



- Generative story and Bayes net



$$p(W_3 \mid U_1=T, U_2=T, U_3=F)$$

- Assumptions
- Joint distribution

Image: http://ai.berkeley.edu/

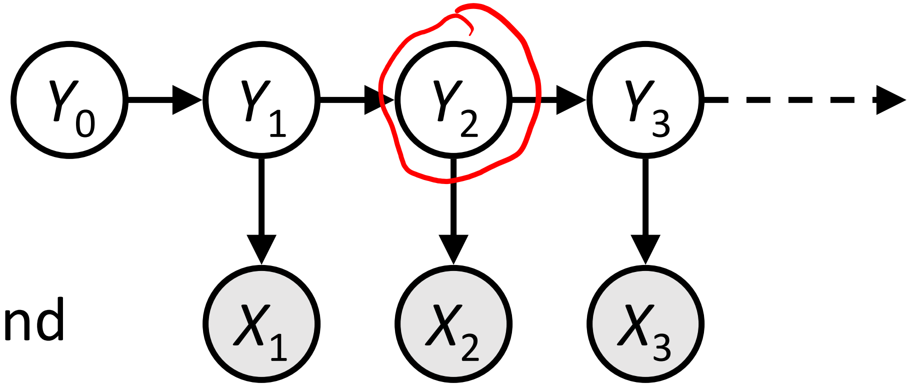
Hidden Markov Models

Image: http://ai.berkeley.edu/

Hidden Markov Models

Usually the true state is not observed directly

Hidden Markov models (HMMs)



- Underlying Markov chain over states Y
- You observe evidence X at each time step
- Y_t is a single discrete variable; X_t may be continuous and may consist of several variables

HMM conditional independence

- Past Y and future Y independent given the present Y_t
- Past X and future X independent given the present Y_t
- Past X and future Y independent given the present Y_t
- Past Y and future X independent given the present Y_t

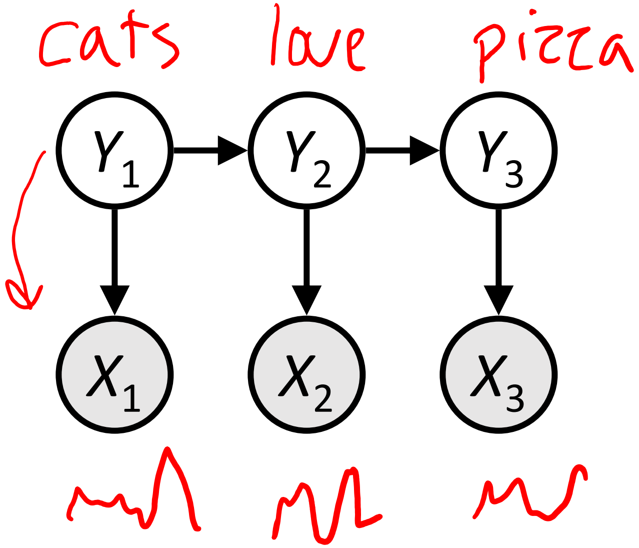
Past $X \perp\!\!\!\perp$ current $X_t \mid Y_t$

Future $X \perp\!\!\!\perp X_t \mid Y_t$

Generative Stories and Bayes Nets

Speech recognition

- Generative story and Bayes net



Y_t : word$_t$

X_t : audio for word$_t$

- Assumptions: HMM conditional independence assumptions

- Joint distribution: $P(Y_1, \cdots, Y_T, X_1, \cdots X_T) = P(Y_1)\prod P(Y_{t+1} \mid Y_t)\prod P(X_t \mid Y_t)$

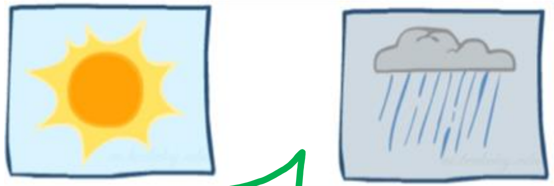
Image: http://ai.berkeley.edu/

Example: Weather HMM

An HMM is defined by:

- Initial distribution: $P(W_0)$
- Transition model: $P(W_t \mid W_{t-1})$
- Emission model: $P(U_t \mid W_t)$

W_{t-1}	$P(W_t \mid W_{t-1})$	
	sun	rain
sun	0.9	0.1
rain	0.3	0.7



W_t	$P(U_t \mid W_t)$	
	true	false
sun	0.2	0.8
rain	0.9	0.1

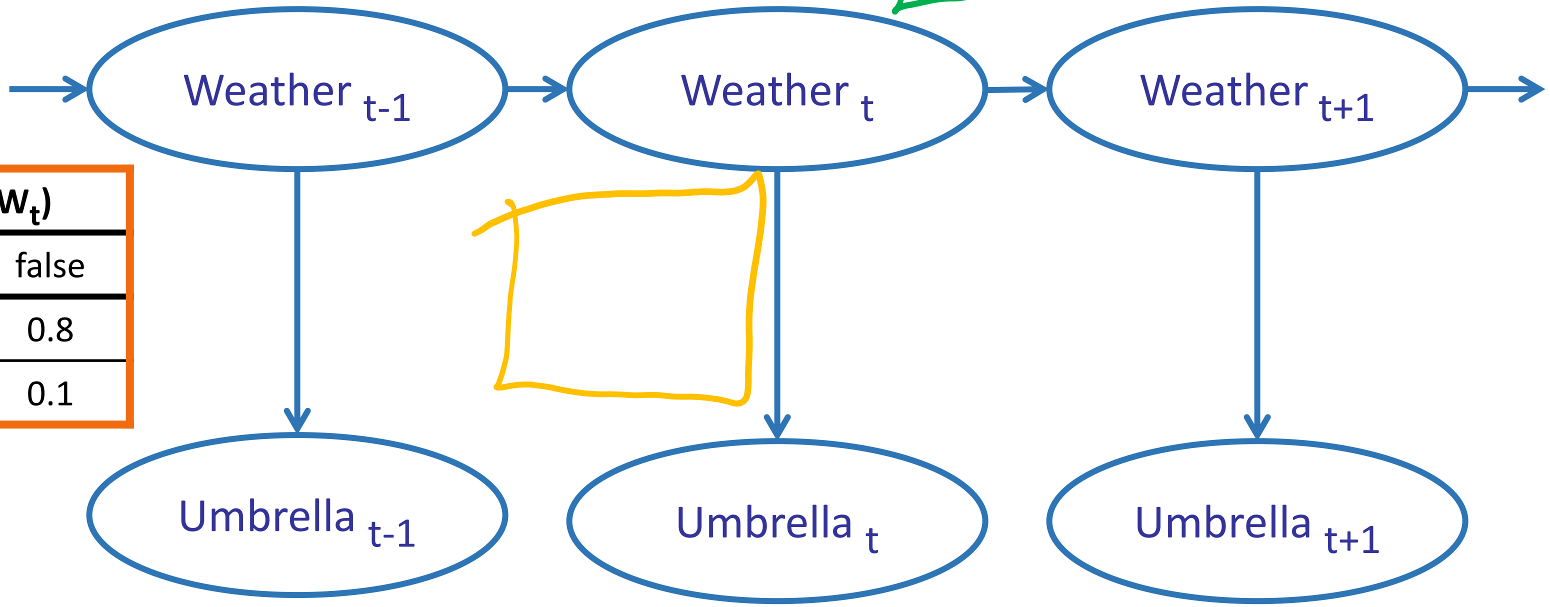
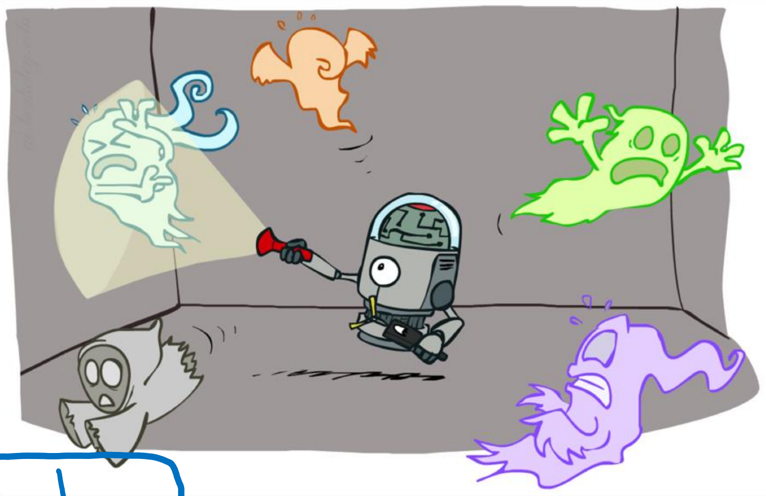


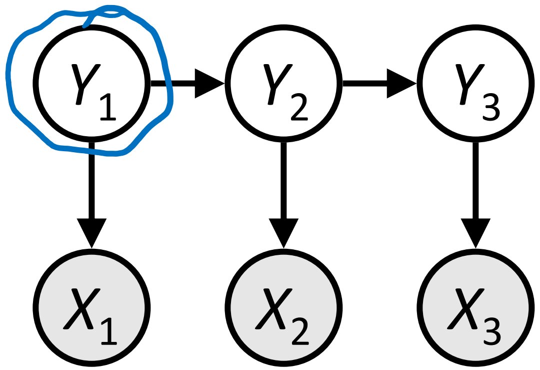
Image: http://ai.berkeley.edu/

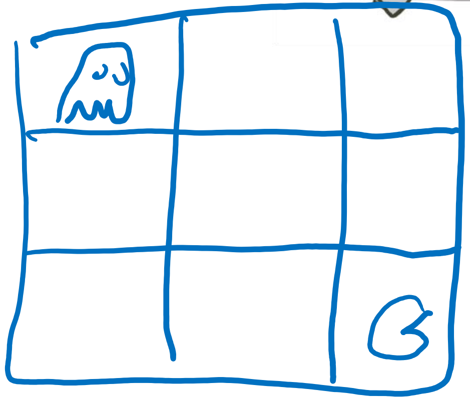
Generative Stories and Bayes Nets



Tracking: Ghostbusting

- Generative story and Bayes net





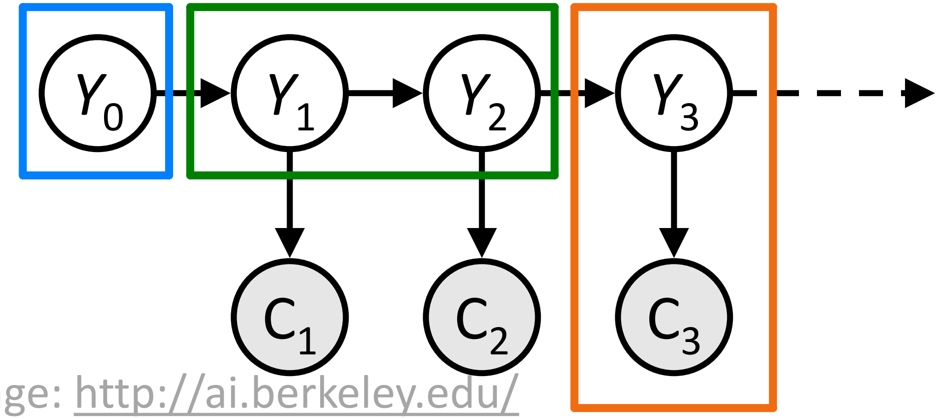
- Assumptions: HMM conditional independence assumptions

- Joint distribution: $P(Y_1, \cdots, Y_T, X_1, \cdots X_T) = P(Y_1) \prod P(Y_{t+1} \mid Y_t) \prod P(X_t \mid Y_t)$

Image: http://ai.berkeley.edu/

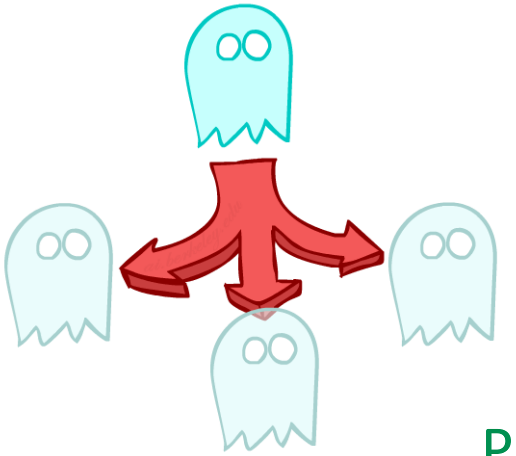
Example: Ghostbusters HMM

- State: location of moving ghost
- Observations: Color recorded by ghost sensor at clicked squares
- $P(Y_0)$ = uniform
- $P(Y_t \mid Y_{t-1})$ = usually move clockwise, but sometimes move randomly or stay in place
- $P(C_t \mid Y_t)$ = sensor model: red means close, green means far away.



1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

$P(Y_1)$



1/6	1/6	1/2
0	1/6	0
0	0	0

$P(Y_2 \mid Y_1=(2,3))$

$P(Y_2 \mid Y_1)$

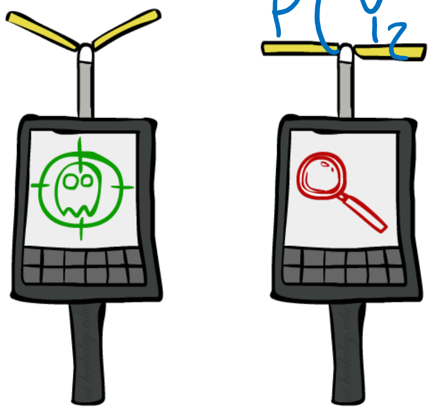
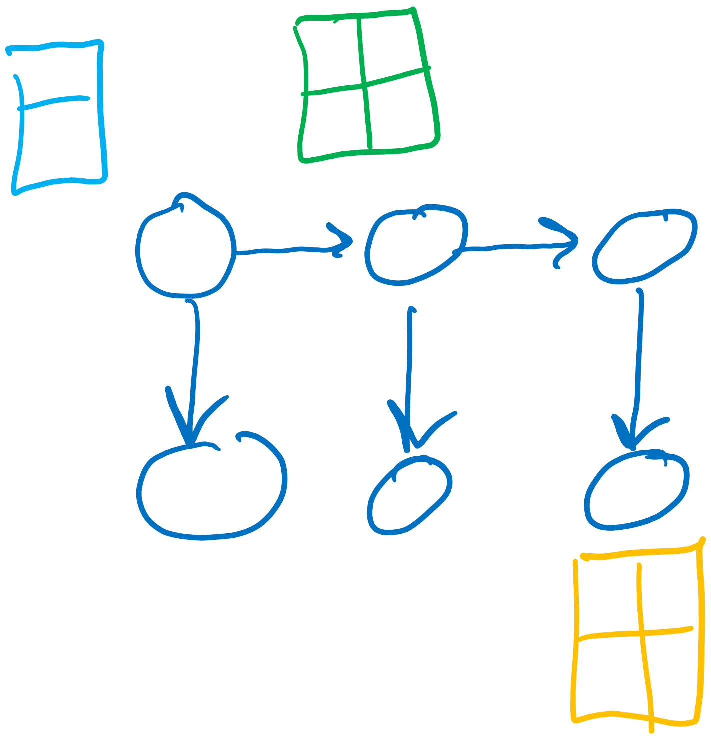


Image: http://ai.berkeley.edu/

Outline

1. Probability primer
2. Generative stories and Bayes nets
   - Bayes nets definition
   - Naïve Bayes
   - Markov chains
   - Hidden Markov models
3. Learning HMM parameters
   - MLE for categorical distribution
4. Inference in Bayes Nets and HMMs



$$p(Y_3 \mid c_1, c_2, c_3)$$

Piazza Poll 1

Assume Y is a discrete random variable taking on 7 distinct values. For example choice of fruit on a given day:

$$y \in \{apple, banana, orange, strawberry, watermelon, pear, grape\}$$

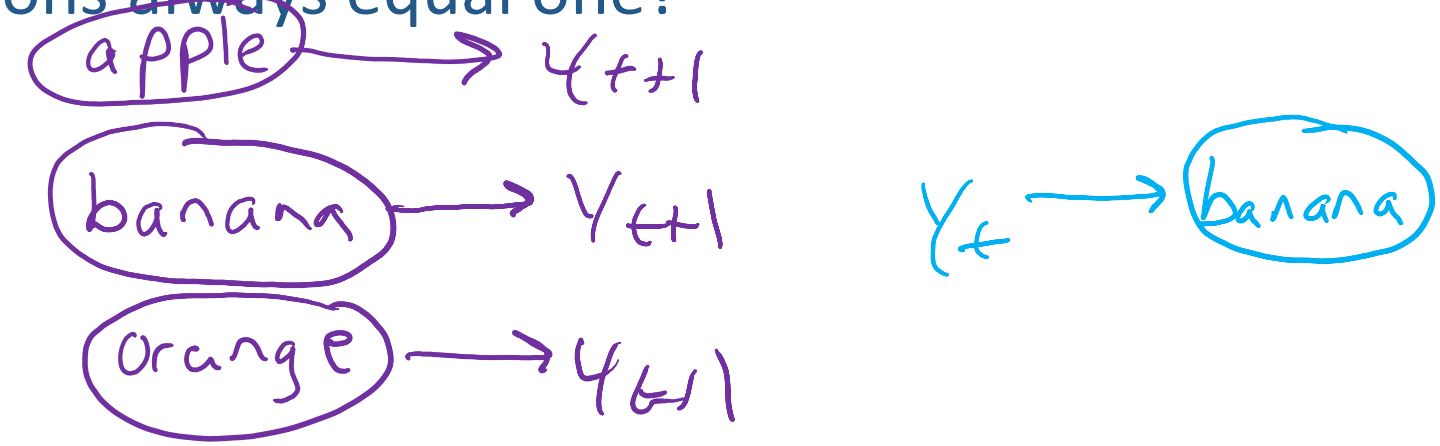
How many entries are in the conditional probability table $P(Y_{t+1} \mid Y_t)$?

A. 7
B. 14
C. 49
D. 2^7
E. 7!

Piazza Poll 2

Which of the following expressions always equal one?

Select ALL that apply



A. $P(y_{t+1} \mid y_t)$

B. $\sum_{y_t \in \mathcal{Y}} P(y_{t+1} \mid y_t)$

C. $\sum_{y_{t+1} \in \mathcal{Y}} P(y_{t+1} \mid y_t)$

D. $\sum_{y_{t+1} \in \mathcal{Y}} \sum_{y_t \in \mathcal{Y}} P(y_{t+1} \mid y_t)$ $= 7$

If it's helpful, consider the fruit example:

$y \in \{apple, banana, orange, strawberry, watermelon, pear, grape\}$

Piazza Poll 3

s, s, s, r, r, r, s, r, r, s, s

How do could we estimate $P(Y_{t+1} = b \mid Y_t = a)$ from data?

A. #(start in state a, end in state b) / #(start in state a)

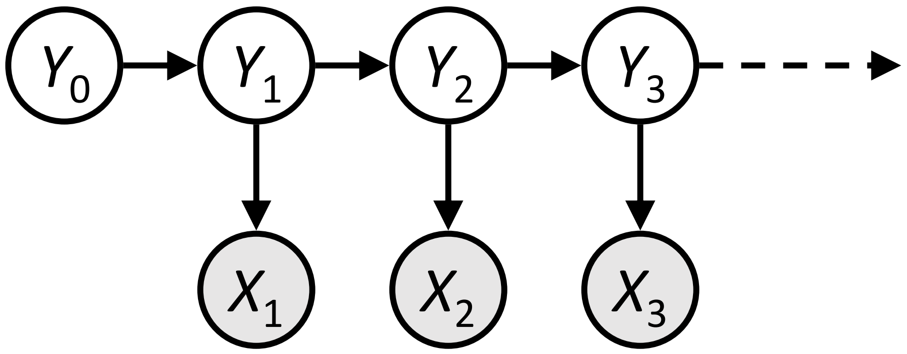
B. #(start in state a, end in state b) / #(end in state b)

C. I have no idea

#( ) notation is the count of occurrences

HMM MLE

Estimate probabilities of categorical distributions



Parameters for:

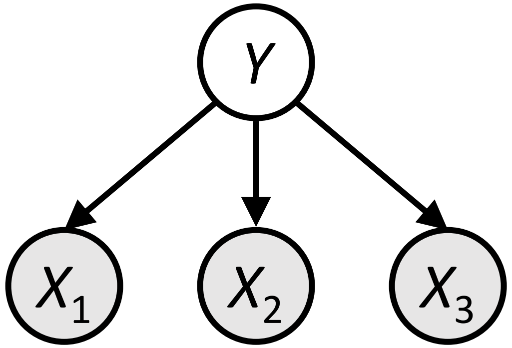
Initial: $P(Y_0)$ ← $|\mathcal{Y}|$

Transition: $P(Y_t \mid Y_{t-1})$ ← $|\mathcal{Y}|^2$

Emission: $P(X_t \mid Y_t)$ ← $|\mathcal{Y}||\mathcal{X}|$

Reminder: Naïve Bayes MLE

SPAM: Bag of words, naïve Bayes



Parameters for:

Class prior: $P(Y)$ ϕ

Class conditional: $P(X_m \mid Y)$

$\theta_{m,y=0}$

$\theta_{m,y=1}$

Reminder: Naïve Bayes MLE

$$\mathcal{D} = \{y^{(n)}, \boldsymbol{x}^{(n)}\}_{n=1}^{N}$$
$$y^{(n)} \in \{0,1\}$$
$$\boldsymbol{x}^{(n)} \in \{0,1\}^M$$
$$\phi \in [0,1]$$
$$\boldsymbol{\Theta} \in [0,1]^{Mx2}$$

$$L(\phi, \boldsymbol{\Theta}) = p(\mathcal{D} \mid \phi, \boldsymbol{\Theta})$$

$$= \prod_{n=1}^{N} p\left(\mathcal{D}^{(n)} \mid \phi, \boldsymbol{\Theta}\right)$$ i.i.d assumption

$$= \prod_{n=1}^{N} p\left(y^{(n)}, \boldsymbol{x}^{(n)} \mid \phi, \boldsymbol{\Theta}\right)$$

$$= \prod_{n=1}^{N} p\left(y^{(n)} \mid \phi\right) p\left(\boldsymbol{x}^{(n)} \mid y^{(n)}, \boldsymbol{\Theta}\right)$$ Generative model

$$= \prod_{n=1}^{N} p\left(y^{(n)} \mid \phi\right) p\left(x_1^{(n)}, x_2^{(n)}, \dots, x_M^{(n)} \mid y^{(n)}, \boldsymbol{\Theta}\right)$$

$$= \prod_{n=1}^{N} p\left(y^{(n)} \mid \phi\right) \prod_{m=1}^{M} p\left(x_m^{(n)} \mid y^{(n)}, \theta_{m,y}\right)$$ Naïve Bayes

$$= \prod_{n=1}^{N} \phi^{y^{(n)}} (1-\phi)^{1-y^{(n)}} \prod_{m=1}^{M} \theta_{m,1}^{\mathbb{I}\left(y^{(n)}=1 \wedge x_m^{(n)}=1\right)} \left(1-\theta_{m,1}\right)^{\mathbb{I}\left(y^{(n)}=1 \wedge x_m^{(n)}=0\right)} \theta_{m,0}^{\mathbb{I}\left(y^{(n)}=0 \wedge x_m^{(n)}=1\right)} \left(1-\theta_{m,0}\right)^{\mathbb{I}\left(y^{(n)}=0 \wedge x_m^{(n)}=0\right)}$$

$$= \phi^{N_{y=1}} (1-\phi)^{N_{y=0}} \prod_{m=1}^{M} \theta_{m,1}^{N_{y=1,x_m=1}} \left(1-\theta_{m,1}\right)^{N_{y=1,x_m=0}} \theta_{m,0}^{N_{y=0,x_m=1}} \left(1-\theta_{m,0}\right)^{N_{y=0,x_m=0}}$$

Reminder: Naïve Bayes MLE

$$L(\phi, \boldsymbol{\Theta}) = p(\mathcal{D} \mid \phi, \boldsymbol{\Theta})$$

$$= \phi^{N_{y=1}} (1-\phi)^{N_{y=0}} \prod_{m=1}^{M} \theta_{m,1}^{N_{y=1,x_m=1}} \left(1-\theta_{m,1}\right)^{N_{y=1,x_m=0}} \theta_{m,0}^{N_{y=0,x_m=1}} \left(1-\theta_{m,0}\right)^{N_{y=0,x_m=0}}$$

$$\ell(\phi, \boldsymbol{\Theta}) = \log p(\mathcal{D} \mid \phi, \boldsymbol{\Theta})$$

$$= \mathrm{N}_{\mathrm{y}=1} \log \phi + \mathrm{N}_{\mathrm{y}=0} \log (1-\phi)$$

$$+ \sum_{m=1}^{M} N_{y=1,x_m=1} \log \theta_{m,1} + N_{y=1,x_m=0} \log \left(1-\theta_{m,1}\right)$$

$$+ \sum_{m=1}^{M} N_{y=0,x_m=1} \log \theta_{m,0} + N_{y=0,x_m=0} \log \left(1-\theta_{m,0}\right)$$

Optimization breaks down for each parameter:

- Set $\frac{\partial \ell}{\partial \phi}$ equal to zero and solve: $\phi = \frac{N_{y=1}}{N_{y=1}+N_{y=0}} = \frac{N_{y=1}}{N}$
- Set $\frac{\partial \ell}{\partial \theta_{m,1}}$ equal to zero and solve: $\theta_{m,1} = \frac{N_{y=1,x_m=1}}{N_{y=1,x_m=1}+N_{y=1,x_m=0}} = \frac{N_{y=1,x_m=1}}{N_{y=1}}$

HMM MLE

Categorical distributions

- Initial: $Y_0 \sim Categorical(\boldsymbol{\phi}_{initial}), \quad y_0 \in \{1 \ldots J\}$
- Transition (given $Y_t = y_t$): $Y_{t+1} \sim Categorical\left(\boldsymbol{\phi}_{trans,y_t}\right), \quad y_{t+1} \in \{1 \ldots J\}$
- Emission (given $Y_t = y_t$): $X_t \sim Categorical\left(\boldsymbol{\phi}_{emiss,y_t}\right), \quad x_t \in \{1 \ldots K\}$

$P(Y_0)$ $P(Y_{t+1} \mid y_t)$ $P(X_t \mid y_t)$

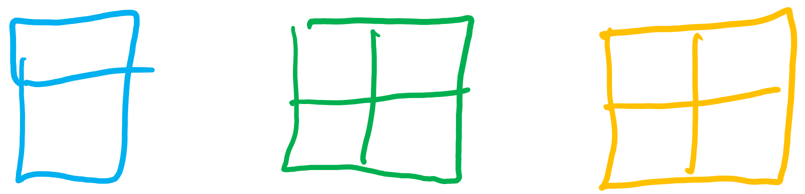
Optimization breaks down for each parameter:

(With Lagrange multiplier trick on constraint that each $\boldsymbol{\phi}$ vector sum to 1, $\sum_i \phi_i = 1$)

- Set $\frac{\partial \ell}{\partial \phi_{initial,j}}$ equal to zero and solve: $\quad \phi_{initial,j} = \frac{\#(Y_0{=}j)}{\sum_{i=1}^{J} \#(Y_0{=}i)} = \frac{\#(Y_0{=}j)}{N}$
- Set $\frac{\partial \ell}{\partial \phi_{trans,y_t,j}}$ equal to zero and solve: $\quad \phi_{trans,y_t,j} = \frac{\#(Y_{t+1}{=}j, Y_t{=}y_t)}{\sum_{i=1}^{J} \#(Y_{t+1}{=}i, Y_t{=}y_t)} = \frac{\#(Y_{t+1}{=}j, Y_t{=}y_t)}{\#(Y_t{=}y_t)}$
- Set $\frac{\partial \ell}{\partial \phi_{emiss,y_t,k}}$ equal to zero and solve: $\quad \phi_{emiss,y_t,k} = \frac{\#(X_t{=}k, Y_t{=}y_t)}{\sum_{i=1}^{K} \#(X_t{=}i, Y_t{=}y_t)} = \frac{\#(X_t{=}k, Y_t{=}y_t)}{\#(Y_t{=}y_t)}$

Outline

1. Probability primer
2. Generative stories and Bayes nets
   - Bayes nets definition
   - Naïve Bayes
   - Markov chains
   - Hidden Markov models
3. Learning HMM parameters
   - MLE for categorical distribution
4. Inference in Bayes Nets and HMMs



Outline

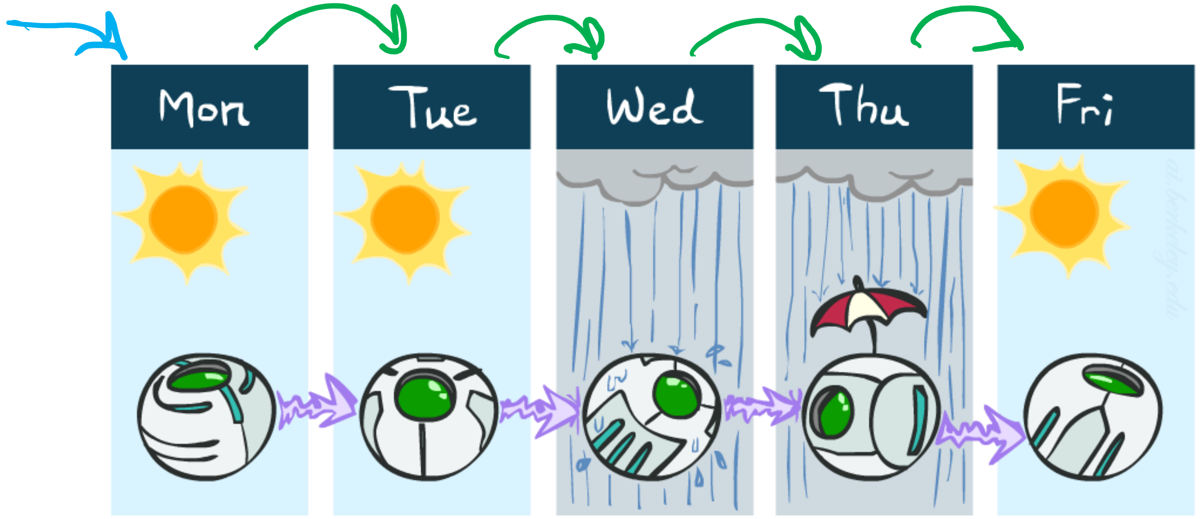
1. Probability primer
2. Generative stories and Bayes nets
   - Bayes nets definition
   - Naïve Bayes
   - Markov chains
   - Hidden Markov models
3. Learning HMM parameters
   - MLE for categorical distribution
4. Inference in Bayes Nets and HMMs

Example: Markov Chain Weather

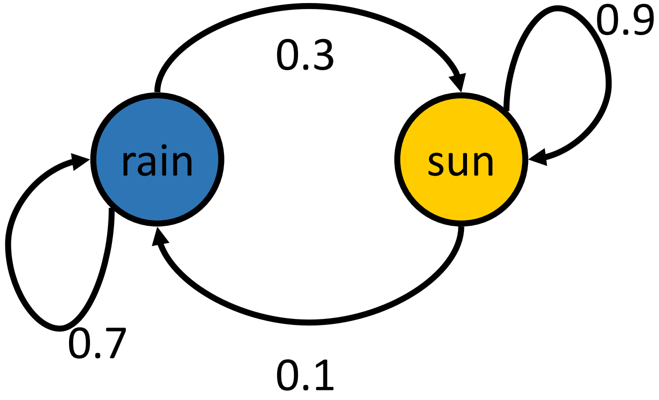
States: Y = {rain, sun}

- Initial distribution: 1.0 sun (0.0 rain)
- Conditional probability table (CPT) $P(Y_t \mid Y_{t-1})$:

Y_{t-1}	Y_t	$P(Y_t \mid Y_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7



Two other ways of representing the same CPT



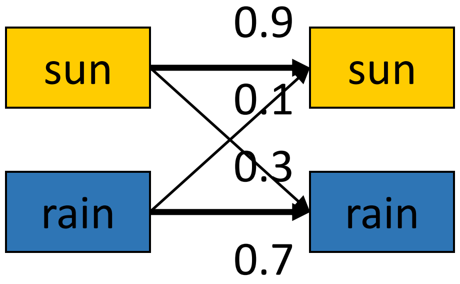


Image: http://ai.berkeley.edu/

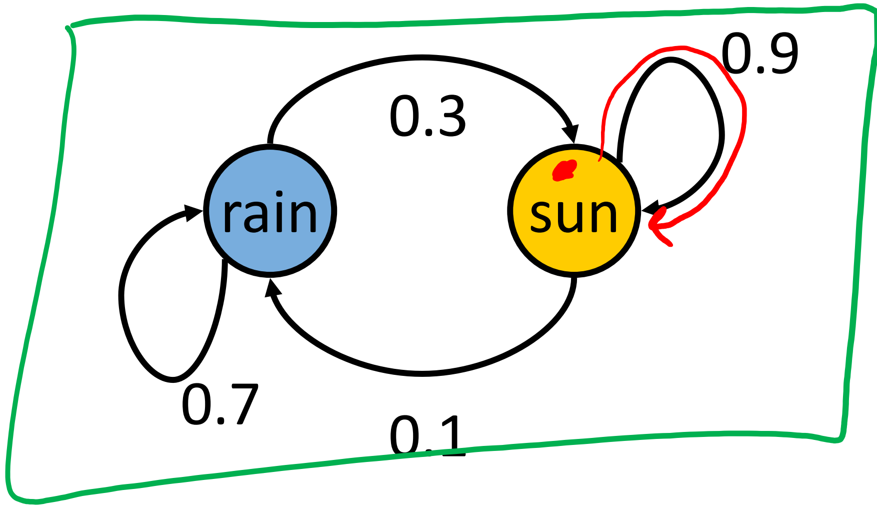
Example: Markov Chain Weather

Initial distribution: $P(Y_1 = sun) = 1.0$

$P(Y_1 = rain) = 0.0$

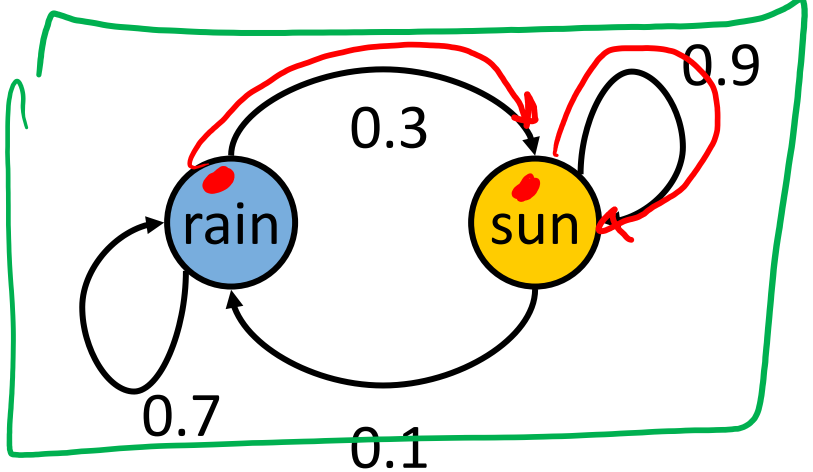
What is the probability distribution after one step?

$P(Y_2 = sun) = ?$



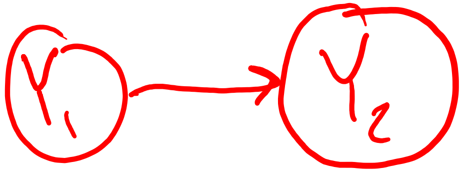
Example: Markov Chain Weather

Initial distribution: $P(Y_1 = sun) = 1.0$



What is the probability distribution after one step?

$P(Y_2 = sun) = ?$



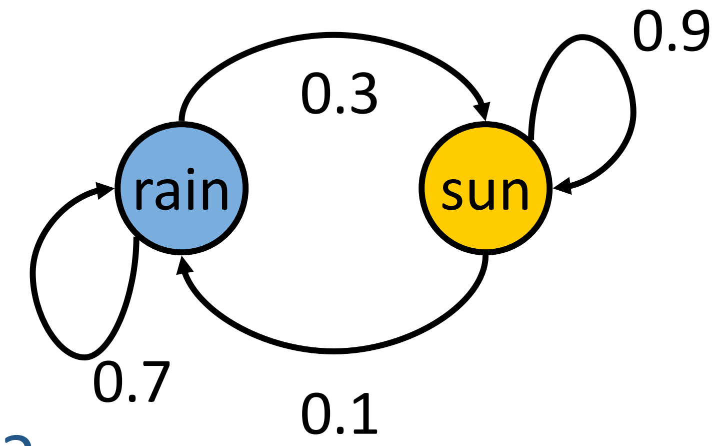
$$P(Y_2 = sun) = \sum_{y_1} P(Y_1 = y_1, Y_2 = sun)$$

$$= \sum_{y_1} P(Y_2 = sun \mid Y_1 = y_1) P(Y_1 = y_1)$$

$$= P(Y_2 = sun \mid Y_1 = sun) P(Y_1 = sun) + P(Y_2 = sun \mid Y_1 = rain) P(Y_1 = rain)$$

$$= 0.9 \cdot 1.0 + 0.3 \cdot 0.0 = 0.9$$

Piazza Poll 4



Initial distribution: $P(Y_2 = sun) = 0.9$

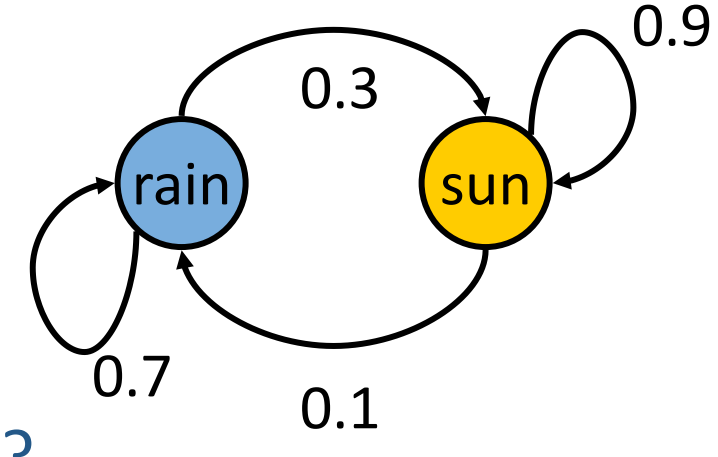
$P(Y_2 = rain) = 0.1$

What is the probability distribution after the next step?

$P(Y_3 = sun) = ?$

A) 0.81
B) 0.84
C) 0.9
D) 1.0
E) 1.2

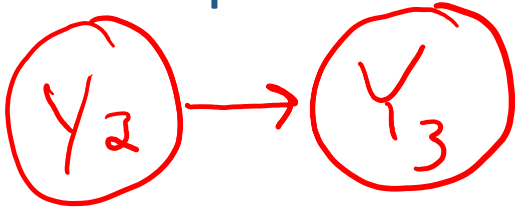
Piazza Poll 4



Initial distribution: $P(Y_2 = sun) \ = 0.9$

What is the probability distribution after the next step?

$P(Y_3 = sun) \ = ?$



A) 0.81

B) 0.84

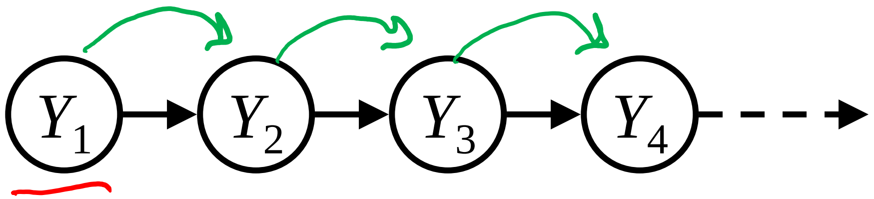
C) 0.9

D) 1.0

E) 1.2

$$
\begin{aligned}
P(Y_3 = sun) &= \sum_{y_2} P(Y_2 = y_2, Y_3 = sun) \\
&= \sum_{y_2} P(Y_3 = sun \mid Y_2 = y_2)P(Y_2 = y_2) \\
&= P(Y_3 = sun \mid Y_2 = sun)P(Y_2 = sun) + \\
&\quad P(Y_3 = sun \mid Y_2 = rain)P(Y_2 = rain) \\
&= 0.9 \cdot 0.9 + 0.3 \cdot 0.1 \\
&= 0.81 + 0.3 \\
&= \underline{0.84}
\end{aligned}
$$

Markov Chain Inference



If you know the transition probabilities, $P(Y_t \mid Y_{t-1})$, and you know $P(Y_4)$, write an equation to compute $P(Y_5)$.

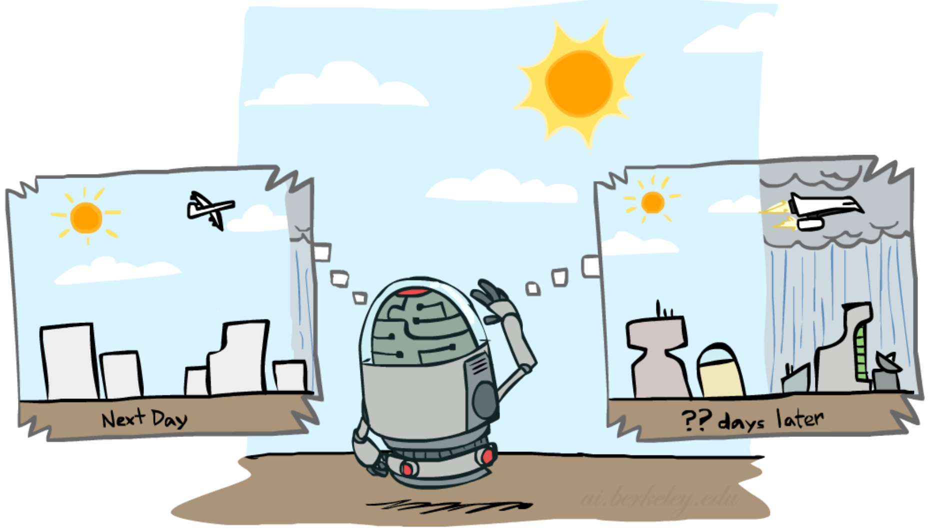
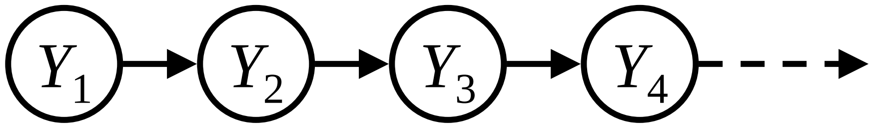


Image: http://ai.berkeley.edu/

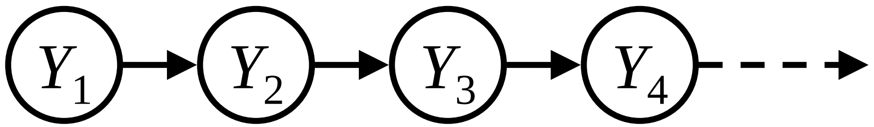
Markov Chain Inference



If you know the transition probabilities, $P(Y_t \mid Y_{t-1})$, and you know $P(Y_4)$, write an equation to compute $P(Y_5)$.

$$P(Y_5) = \sum_{y_4} P(y_4, Y_5)$$
$$= \sum_{y_4} P(Y_5 \mid y_4) P(y_4)$$

Markov Chain Inference



If you know the transition probabilities, $P(Y_t \mid Y_{t-1})$, and you know $P(Y_4)$, write an equation to compute $P(Y_5)$.

$Y_1 \to Y_2 \to Y_3 \to Y_4 \to Y_5$

$$P(Y_5) = \sum_{y_1,y_2,y_3,y_4} P(y_1, y_2, y_3, y_4, Y_5)$$

$$= \sum_{y_1,y_2,y_3,y_4} P(Y_5 \mid y_4) P(y_4 \mid y_3) P(y_3 \mid y_2) P(y_2 \mid y_1) P(y_1)$$

$$= \sum_{y_4} P(Y_5 \mid y_4) \sum_{y_1,y_2,y_3} P(y_4 \mid y_3) P(y_3 \mid y_2) P(y_2 \mid y_1) P(y_1)$$

$$= \sum_{y_4} P(Y_5 \mid y_4) \sum_{y_1,y_2,y_3} P(y_1, y_2, y_3, y_4)$$

$$= \sum_{y_4} P(Y_5 \mid y_4) P(y_4)$$

Example: Markov Chain Weather

States {rain, sun}

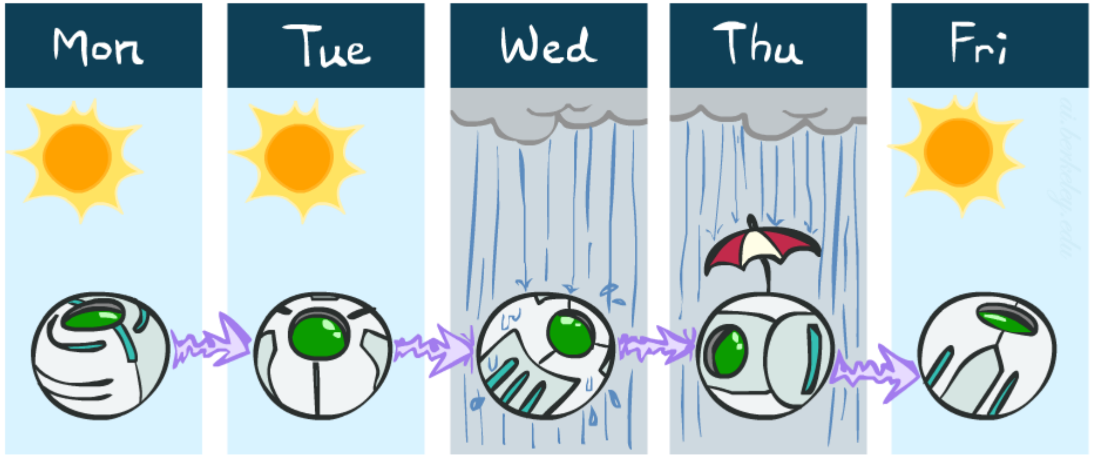
- Initial distribution $P(Y_0)$

$P(Y_0)$	
sun	rain
0.5	0.5

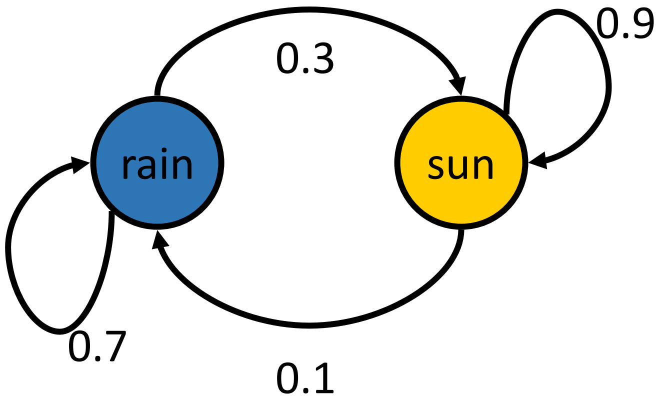
- Transition model $P(X_t \mid X_{t-1})$

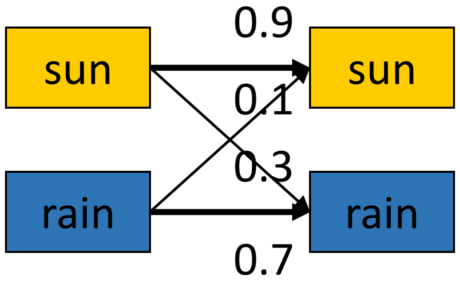
Y_{t-1}	$P(Y_t \mid Y_{t-1})$	
	sun	rain
sun	0.9	0.1
rain	0.3	0.7

Image: http://ai.berkeley.edu/



Two other ways of representing the same CPT

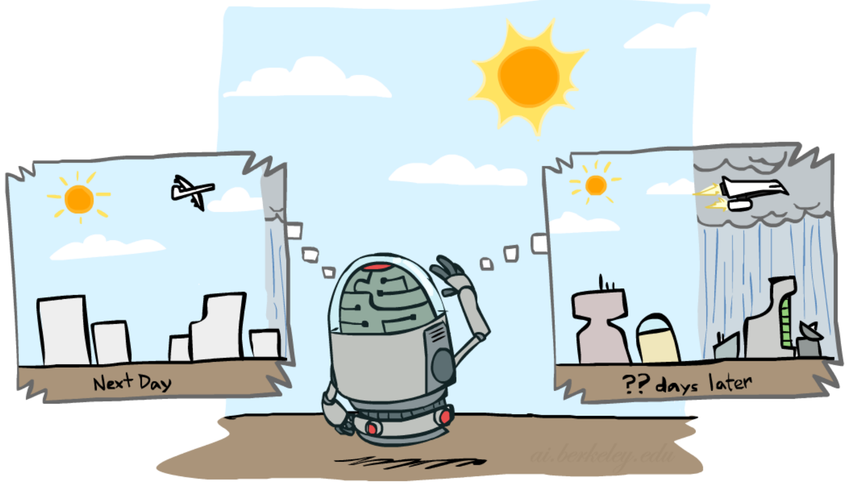




Weather prediction

Time 0: $P(Y_0) =<0.5,0.5>$

Y_{t-1}	$P(Y_t \mid Y_{t-1})$	
	sun	rain
sun	0.9	0.1
rain	0.3	0.7



What is the weather like at time 1?

$$P(Y_1) = \sum_{y_0} P(Y_0=y_0, Y_1)$$

$$= \sum_{y_0} P(Y_1 \mid Y_0=y_0)\, P(Y_0=y_0)$$

$$= 0.5<0.9,0.1> + 0.5<0.3,0.7>$$

$$= <0.6,0.4>$$

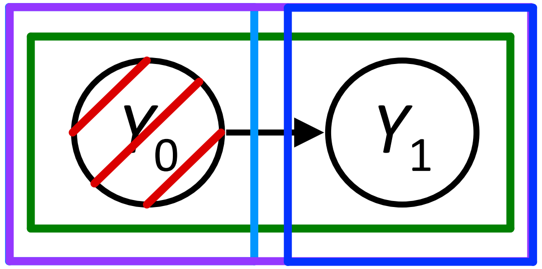
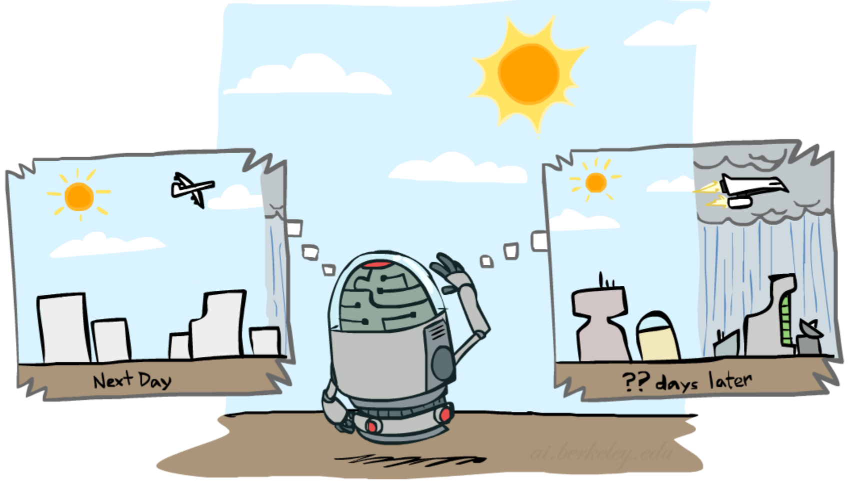


Image: http://ai.berkeley.edu/

Weather prediction, contd.

Time 1: $P(Y_1) = \langle 0.6, 0.4 \rangle$

Y_{t-1}	$P(Y_t \mid Y_{t-1})$	
	sun	rain
sun	0.9	0.1
rain	0.3	0.7



What is the weather like at time 2?

$$P(Y_2) = \sum_{y_1} P(Y_1 = y_1, Y_2)$$

$$= \sum_{y_1} P(Y_2 \mid Y_1 = y_1)\, P(Y_1 = y_1)$$

$$= 0.6\langle 0.9, 0.1\rangle + 0.4\langle 0.3, 0.7\rangle$$

$$= \langle 0.66, 0.34 \rangle$$

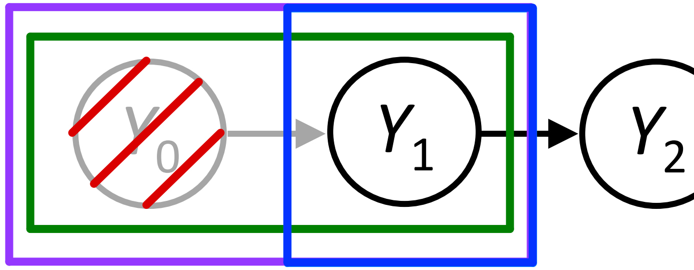
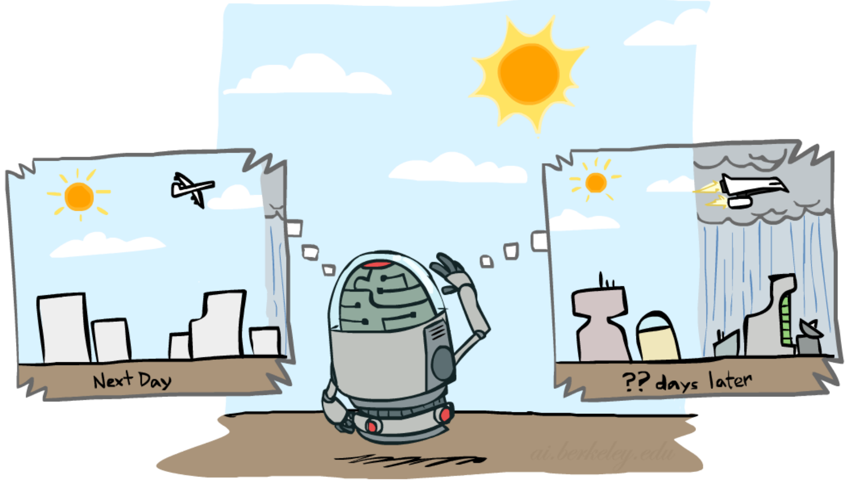


Image: http://ai.berkeley.edu/

Weather prediction, contd.

Time 2: $P(Y_2)$ =<0.66,0.34>

Y_{t-1}	$P(Y_t \mid Y_{t-1})$	
	sun	rain
sun	0.9	0.1
rain	0.3	0.7



What is the weather like at time 3?

$$P(Y_3) = \sum_{y_2} P(Y_2=y_2, Y_3)$$
$$= \sum_{y_2} P(Y_3 \mid Y_2=y_2)\, P(Y_2=y_2)$$
$$= 0.66\langle 0.9, 0.1\rangle + 0.34\langle 0.3, 0.7\rangle$$
$$= \langle 0.696, 0.304\rangle$$

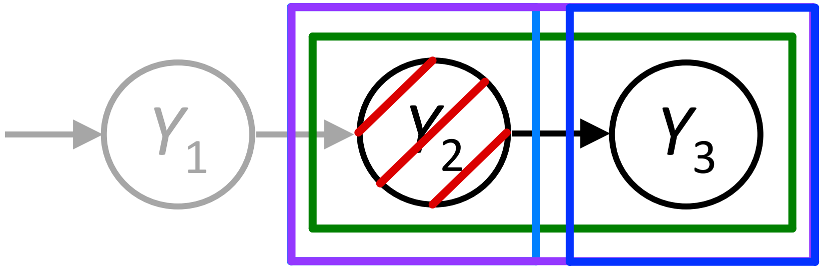


Image: http://ai.berkeley.edu/

Forward algorithm (simple form)

What is the state at time t?

$$P(Y_t) = \sum_{y_{t-1}} P(Y_{t-1}=y_{t-1}, Y_t)$$
$$= \sum_{y_{t-1}} P(Y_t \mid Y_{t-1}=y_{t-1})\, P(Y_{t-1}=y_{t-1})$$

Iterate this update starting at t=1

Inference: Hidden Markov Models

Image: http://ai.berkeley.edu/

HMM as Probability Model

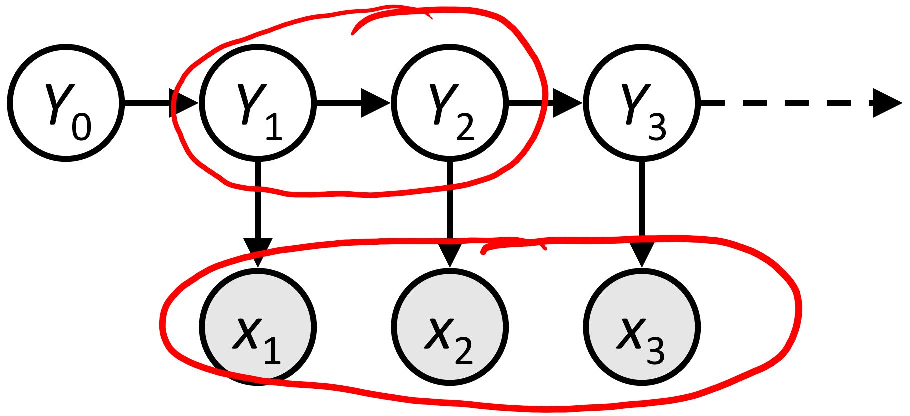
- Joint distribution for Markov model:

$$P(Y_0,\dots, Y_T) = P(Y_0) \prod_{t=1:T} P(Y_t \mid Y_{t-1})$$

- Joint distribution for hidden Markov model:

$$P(Y_0, Y_1,X_1, \dots, Y_T,X_T) = P(Y_0) \prod_{t=1:T} P(Y_t \mid Y_{t-1})\, P(X_t \mid Y_t)$$

- Future states are independent of the past given the present
- Current evidence is independent of everything else given the current state
- Are evidence variables independent of each other?



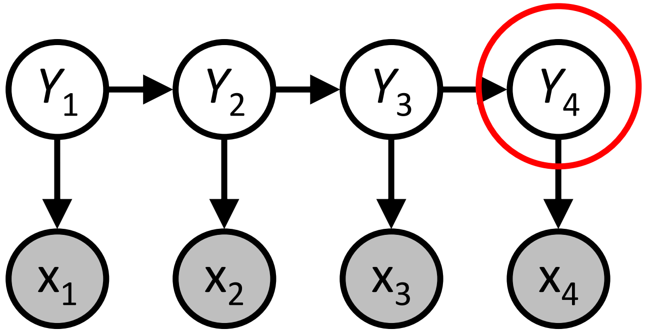
Notation alert!

Useful notation: $Y_{a:b} = Y_a, Y_{a+1}, \dots, Y_b$

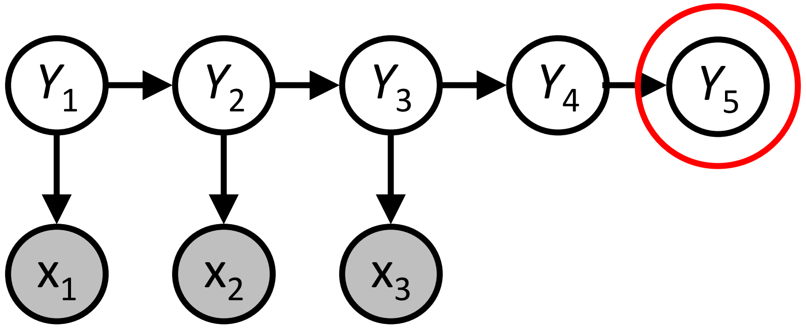
For example: $P(Y_{1:2} \mid x_{1:3}) = P(Y_1, Y_2, \mid x_1, x_2, x_3)$

HMM Queries

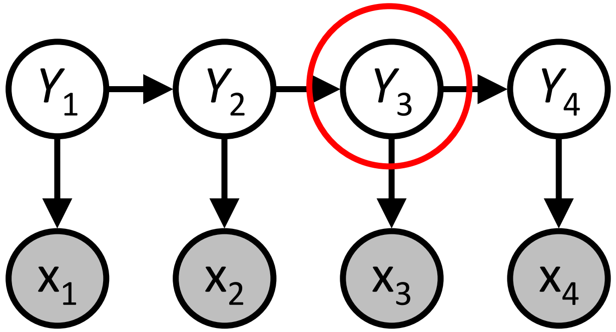
Filtering: $P(Y_t | x_{1:t})$



Prediction: $P(Y_{t+k} | x_{1:t})$



Smoothing: $P(Y_k | x_{1:t})$, $k<t$



Explanation: $P(Y_{1:t} | x_{1:t})$