http://www.youtube.com/watch?v=jPGgl5VH5go

Hidden Markov Models

David Kauchak, CS151, Fall 2010

Admin

- Final project
  - Experimentally evaluate some idea/method we've discussed (or will discus) in class
  - Must work in pairs (come talk to me if this causes a problem)
  - Schedule:
    - Literature review (this week)
    - Proposal
    - Status report(s)
    - Paper draft
    - Peer reviews
    - Final paper
    - Presentations
- Homework 6 out soon
- Last assignment next week on HMMs
- Midterm
  - average: 45
  - high: 56.5 (2nd highest 51)

Bayes rule: a refresher

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

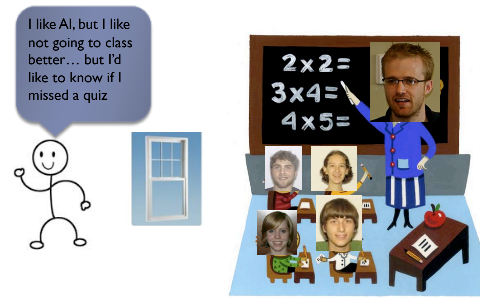
$$P(A \mid B,C) = \frac{P(B,C \mid A)P(A)}{P(B,C)} = \frac{P(B \mid A,C)P(A \mid C)}{P(B \mid C)} = \frac{P(C \mid A,B)P(A \mid B)}{P(C \mid B)}$$

We'll see this version a lot today

Predicting the Quiz

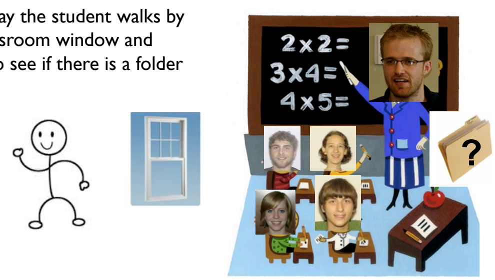
- I give quizzes in CS51 on written problems
- Often, I print out the problem and hand out the quiz at the beginning
  - One class, I gave a quiz, but had the students use their own paper
  - Sometimes I bring graded quizzes or exams to hand back
- In both situations, students have mispredicted quizzes

The setup



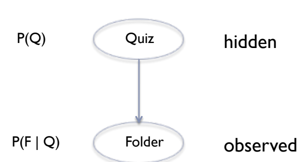
The setup

Every day the student walks by the classroom window and looks to see if there is a folder

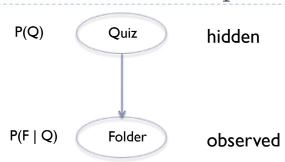


Bayesian model for this process?

- What are the variables?
  - Folder/no-folder
  - Quiz/no-quiz
- How are they related?

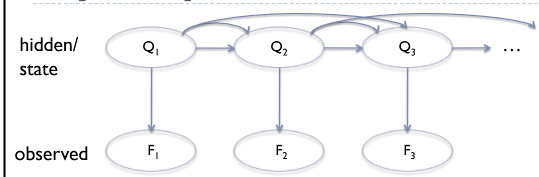


Bayesian model for this process?



Any other information?
What about from day to day?

Sequential process!



What would these conditional probability distributions look like?

$P(Q_1)\ P(F_1 \mid Q_1)$

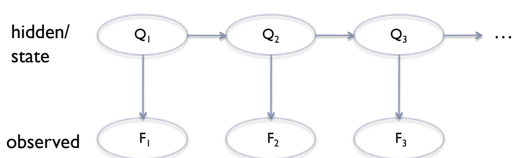
$P(Q_2 \mid Q_1)\ P(F_2 \mid Q_2)$

$P(Q_3 \mid Q_1\ Q_2)\ P(F_3 \mid Q_3)$

$P(Q_i \mid Q_1\ Q_2 \ldots Q_{i-1})\ P(F_i \mid Q_i)$

Any assumption that could reduce the parameter space?

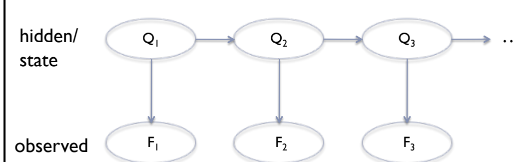
Markov assumption



Markov assumption: the current state (Q) only depends on a fixed number of previous states

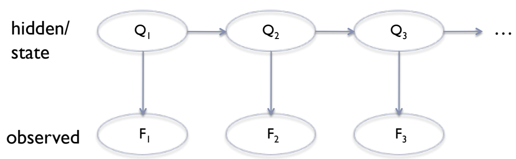
first-order Markov assumption: only the previous state

Hidden Markov Model



Called a Hidden Markov Model (or Dynamic Bayes Net)*
*these two are different, we'll differentiate later

Hidden Markov Models



What are the CPDs?

$P(Q_1)\ P(F_1 \mid Q_1)$

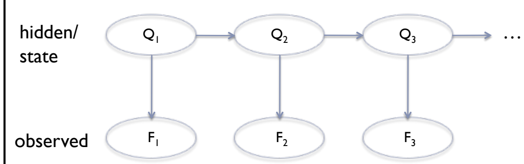
$P(Q_2 \mid Q_1)\ P(F_2 \mid Q_2)$

$P(Q_3 \mid Q_2)\ P(F_3 \mid Q_3)$

$P(Q_i \mid Q_{i-1})\ P(F_i \mid Q_i)$

This is still a lot of parameters! Can we tie any parameters, like we did for multinomial NB?

Hidden Markov Models

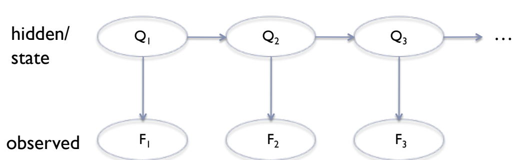


$P(Q_2|Q_1) = P(Q_3|Q_2) = P(Q_4|Q_3) = P(Q_t|Q_{t-1})$

$P(F_1|Q_1) = P(F_2|Q_2) = P(F_3|Q_3) = P(F_t|Q_t)$

Stationary process: value of the state may change but not the underlying distribution

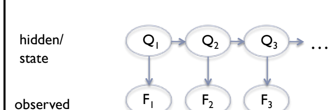
Hidden Markov Models



What are some of the independences/conditional independences?

What questions could we ask?

- Filtering: What is the distribution over Q today?
  - $P(Q_t \mid F_{1:t})$
- Prediction: What is the distribution Q in the future?
  - $P(Q_{t+k} \mid F_{1:t})$ – k days from now
- Smoothing: What is the distribution over Q in the past?
  - $P(Q_k \mid F_{1:t})$ – for $k < t$
- Most likely explanation: What states have we been in?
  - $\text{argmax}_{Q1:t}\ P(Q_{1:t} \mid F_{1:t})$
- Learning
  - Learn CPD tables given observations

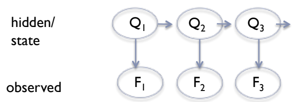


Filtering

Given my observations up to and including today, what is the probability of a Quiz?

$P(Q_t \mid F_{1:t})$

Like to get this defined recursively with respect to $P(Q_{t-1} \mid F_{1:t-1})$



Filtering

Given my observations up to and including today, what is the probability of a Quiz?

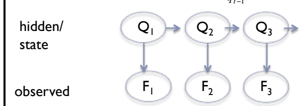
$$P(Q_t \mid F_{1:t}) = P(Q_t \mid F_{1:t-1}, F_t)$$

$$= \alpha P(F_t \mid Q_t, F_{1:t-1}) P(Q_t \mid F_{1:t-1})$$ Bayes' rule

$$= \alpha P(F_t \mid Q_t) P(Q_t \mid F_{1:t-1})$$ Markov model

$$= \alpha P(F_t \mid Q_t) \sum_{q_{t-1}} P(Q_t \mid F_{1:t-1}, q_{t-1}) P(q_{t-1} \mid F_{1:t-1})$$

$$= \alpha P(F_t \mid Q_t) \sum_{q_{t-1}} P(Q_t \mid q_{t-1}) P(q_{t-1} \mid F_{1:t-1})$$ Markov model



Filtering

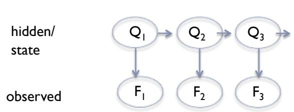
$$P(Q_t \mid F_{1:t}) = \alpha P(F_t \mid Q_t) \sum_{q_{t-1}} P(Q_t \mid q_{t-1}) P(q_{t-1} \mid F_{1:t-1})$$

What is this?

probability of folder given quiz

probability of quiz given quiz yesterday

Can be read from our CPTs



Filtering

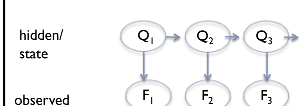
$$P(Q_t \mid F_{1:t}) = \alpha P(F_t \mid Q_t) \sum_{q_{t-1}} P(Q_t \mid q_{t-1}) P(q_{t-1} \mid F_{1:t-1})$$

probability of folder given quiz

probability of quiz given quiz yesterday

Can be read from our CPTs

recursive definition (aka forward message)



Filtering:
What is the probability of a quiz today given no folder for the past two days?

$$P(Q_t \mid F_{1:t}) = \alpha P(F_t \mid Q_t) \sum_{q_{t-1}} P(Q_t \mid q_{t-1}) \cdot message(q_{t-1})$$

Q_{t-1}	Q_t
T	0.1
F	0.5

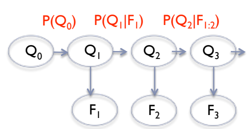
$P(Q_0) = 0.25$

Q_t	F_t
T	0.8
F	0.3

- Start from the beginning and work your way forward
- Each previous answer passes a "message" forward

hidden/ state

observed



Filtering:
What is the probability of a quiz today given no folder for the past two days?

$$P(Q_t \mid F_{1:t}) = \alpha P(F_t \mid Q_t) \sum_{q_{t-1}} P(Q_t \mid q_{t-1}) \cdot message(q_{t-1})$$

Q_{t-1}	Q_t
T	0.1
F	0.5

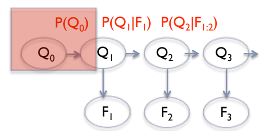
$P(Q_0) = 0.25$

$$P(Q_0) = \begin{bmatrix} 0.25 \\ 0.75 \end{bmatrix}$$

Q_t	F_t
T	0.8
F	0.3

hidden/ state

observed



Filtering:
What is the probability of a quiz today given no folder for the past two days?

$$P(Q_t \mid F_{1:t}) = \alpha P(F_t \mid Q_t) \sum_{q_{t-1}} P(Q_t \mid q_{t-1}) \cdot message(q_{t-1})$$

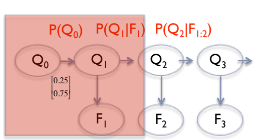
$$P(Q_1 \mid F_1) = \alpha P(\neg f_1 \mid Q_1)\left(P(Q_1 \mid True) \cdot message(True) + P(Q_1 \mid False) \cdot message(False)\right)$$

Q_{t-1}	Q_t
T	0.1
F	0.5

Q_t	F_t
T	0.8
F	0.3

hidden/ state

observed



Filtering:
What is the probability of a quiz today given no folder for the past two days?

$$P(Q_t \mid F_{1:t}) = \alpha P(F_t \mid Q_t) \sum_{q_{t-1}} P(Q_t \mid q_{t-1}) \cdot message(q_{t-1})$$

$$P(Q_1 \mid F_1) = \alpha P(\neg f_1 \mid Q_1)\left(P(Q_1 \mid True) \cdot message(True) + P(Q_1 \mid False) \cdot message(False)\right)$$

$$= \alpha \begin{bmatrix} 0.2 \\ 0.7 \end{bmatrix} \left(\begin{bmatrix} 0.1 \\ 0.9 \end{bmatrix} 0.25 + \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} 0.75 \right)$$

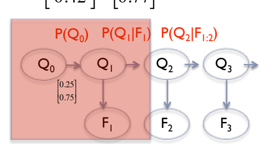
$$= \alpha \begin{bmatrix} 0.125 \\ 0.42 \end{bmatrix} = \begin{bmatrix} 0.23 \\ 0.77 \end{bmatrix}$$

Q_{t-1}	Q_t
T	0.1
F	0.5

Q_t	F_t
T	0.8
F	0.3

hidden/ state

observed



Filtering:
What is the probability of a quiz today given no folder for the past two days?

$$P(Q_t \mid F_{1:t}) = \alpha P(F_t \mid Q_t) \sum_{q_{t-1}} P(Q_t \mid q_{t-1}) \cdot message(q_{t-1})$$

$$P(Q_2 \mid F_{1:2}) = \alpha P(\neg f_2 \mid Q_2)\left(P(Q_2 \mid True) \cdot message(True) + P(Q_2 \mid False) \cdot message(False)\right)$$

$$= \alpha \begin{bmatrix}0.2\\0.7\end{bmatrix}\left(\begin{bmatrix}0.1\\0.9\end{bmatrix}0.23 + \begin{bmatrix}0.5\\0.5\end{bmatrix}0.77\right)$$

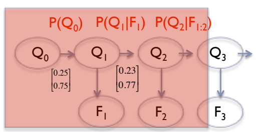
$$= \alpha \begin{bmatrix}0.0816\\0.414\end{bmatrix} = \begin{bmatrix}0.165\\0.835\end{bmatrix}$$

Q_{t-1}	Q_t
T	0.1
F	0.5

Q_t	F_t
T	0.8
F	0.3

hidden/ state

observed



Filtering:
What is the probability of a quiz today given no folder for the past two days?

$$P(Q_t \mid F_{1:t}) = \alpha P(F_t \mid Q_t) \sum_{q_{t-1}} P(Q_t \mid q_{t-1}) \cdot message(q_{t-1})$$

$$P(Q_3 \mid F_{1:3}) = \alpha P(f_3 \mid Q_3)\left(P(Q_3 \mid True) \cdot message(True) + P(Q_3 \mid False) \cdot message(False)\right)$$

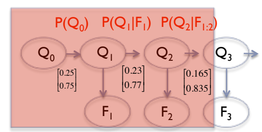
Q_{t-1}	Q_t
T	0.1
F	0.5

Q_t	F_t
T	0.8
F	0.3

We could keep going with more evidence..

hidden/ state

observed

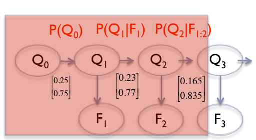


Forward message

$$P(Q_t \mid F_{1:t}) = \alpha P(F_t \mid Q_t) \sum_{q_{t-1}} P(Q_t \mid q_{t-1}) \cdot message(q_{t-1})$$

$$P(Q_3 \mid F_{1:3}) = \alpha P(f_3 \mid Q_3)\left(P(Q_3 \mid True) \cdot message(True) + P(Q_3 \mid False) \cdot message(False)\right)$$

This message is called the "forward" message. Why?



Prediction

- Given our equation for filtering, how can we predict what will happen in the future?
  - $P(Q_{t+k} \mid F_{1:t})$ – k days from now

$$P(Q_t \mid F_{1:t}) = \alpha P(F_t \mid Q_t) \sum_{q_{t-1}} P(Q_t \mid q_{t-1}) P(q_{t-1} \mid F_{1:t-1})$$

probability of folder given quiz

probability of quiz given quiz yesterday

Can be read from our CPTs

recursive (message)

Prediction

- We no longer have evidence, so we only take into account the state probabilities

$$P(Q_{t+k} \mid F_{1:t}) = \alpha \sum_{q_{t+k-1}} P(Q_{t+k} \mid q_{t+k-1}) P(q_{t+k-1} \mid F_{1:t-1})$$

probability of quiz given quiz the previous day

recursive (message)

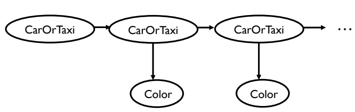
Your turn

- You live on the top floor of a building, giving you a good view of the street from your kitchen window
- Its really cold outside, so you'd rather maximize your time inside and watch for a taxi out the window
- There is a constant stream of vehicles coming in, but you can only make out the color
- You'd like to predict whether or not it's a taxi
- Formulate this problem as an HMM problem

- Step 1: what does the HMM look like?

Your turn

- Problem: Identify the taxi



CT_{t-1}	CT_t	$P(CT_t \mid CT_{t-1})$
C	C	0.5
C	T	0.5
T	C	0.75
T	T	0.25

CT	Col	$P(Col \mid CT_t)$
C	Y	0.1
C	NY	0.9
T	Y	0.75
T	NY	0.25

Your Turn

- You just saw a red vehicle, and now you see a yellow vehicle. What is the probability that the vehicle you see is a taxi? (Assume taxis and cars are equally likely when you start looking)
- What is the probability that the *next* vehicle will be a taxi?

CT_{t-1}	CT_t	$P(CT_t \mid CT_{t-1})$
C	C	0.5
C	T	0.5
T	C	0.75
T	T	0.25

CT	Col	$P(Col \mid CT_t)$
C	Y	0.1
C	NY	0.9
T	Y	0.75
T	NY	0.25

$$P(X_t \mid E_{1:t}) = \alpha P(E_t \mid X_t) \sum_{x_{t-1}} P(X_t \mid x_{t-1}) P(x_{t-1} \mid E_{1:t-1})$$

Filtering: red then yellow, equal prior

CT_{t-1}	CT_t	$P(CT_t \mid CT_{t-1})$
C	C	0.5
C	T	0.5
T	C	0.75
T	T	0.25

CT	Col	$P(Col \mid CT_t)$
C	Y	0.1
C	NY	0.9
T	Y	0.75
T	NY	0.25

$$P(X_i \mid E_i) = \alpha P(f_i \mid X_i)\left(P(X_i \mid C)\cdot message(C) + P(X_i \mid T)\cdot message(T)\right)$$

$$= \alpha P(f_i \mid X_i)\left(\begin{bmatrix}0.5\\0.5\end{bmatrix} message(C) + \begin{bmatrix}0.75\\0.25\end{bmatrix} message(T)\right)$$

First message: [0.50, 0.50] (equal prior)

Filtering: red with equal prior

CT_{t-1}	CT_t	$P(CT_t \mid CT_{t-1})$
C	C	0.5
C	T	0.5
T	C	0.75
T	T	0.25

CT	Col	$P(Col \mid CT_t)$
C	Y	0.1
C	NY	0.9
T	Y	0.75
T	NY	0.25

$$P(X_1 \mid E_1) = \alpha P(NY \mid X_1)\left(P(X_1 \mid C)\cdot message(C) + P(X_1 \mid T)\cdot message(T)\right)$$

$$= \alpha \begin{bmatrix}0.9\\0.25\end{bmatrix}\left(\begin{bmatrix}0.5\\0.5\end{bmatrix}0.5 + \begin{bmatrix}0.75\\0.25\end{bmatrix}0.5\right)$$

$$= \alpha \begin{bmatrix}0.5625\\0.09375\end{bmatrix} = \begin{bmatrix}0.857\\0.143\end{bmatrix}\begin{matrix}car\\taxi\end{matrix}$$

became more probable that it's a car

Filtering: now yellow

CT_{t-1}	CT_t	$P(CT_t \mid CT_{t-1})$
C	C	0.5
C	T	0.5
T	C	0.75
T	T	0.25

CT	Col	$P(Col \mid CT_t)$
C	Y	0.1
C	NY	0.9
T	Y	0.75
T	NY	0.25

$$P(X_i \mid E_i) = \alpha P(Y \mid X_i)\left(P(X_i \mid C)\cdot message(C) + P(X_i \mid T)\cdot message(T)\right)$$

$$= \alpha \begin{bmatrix}0.1\\0.75\end{bmatrix}\left(\begin{bmatrix}0.5\\0.5\end{bmatrix}0.857 + \begin{bmatrix}0.75\\0.25\end{bmatrix}0.143\right)$$

$$= \alpha \begin{bmatrix}0.0536\\0.348\end{bmatrix} = \begin{bmatrix}0.133\\0.867\end{bmatrix}\begin{matrix}car\\taxi\end{matrix}$$

much more likely that it's a taxi

Prediction: a cab next?

CT_{t-1}	CT_t	$P(CT_t \mid CT_{t-1})$
C	C	0.5
C	T	0.5
T	C	0.75
T	T	0.25

CT	Col	$P(Col \mid CT_t)$
C	Y	0.1
C	NY	0.9
T	Y	0.75
T	NY	0.25

$$P(X_i \mid E_i) = \alpha\left(P(X_i \mid C)\cdot message(C) + P(X_i \mid T)\cdot message(T)\right)$$

$$= \alpha\left(\begin{bmatrix}0.5\\0.5\end{bmatrix}0.252 + \begin{bmatrix}0.75\\0.25\end{bmatrix}0.748\right)$$

$$= \begin{bmatrix}0.687\\0.421\end{bmatrix}\begin{matrix}car\\taxi\end{matrix}$$

will slowly move towards a stationary prediction as we look further and further out

Stationary distribution

cars from now	P(car)	P(taxi)
2	0.687	0.313
3	0.57825	0.42175
4	0.6054375	0.3945625
5	0.598640625	0.401359375
6	0.60033984375	0.39966015625
7	0.599915039063	0.400084960938
8	0.600021240234	0.399978759766
9	0.599994689941	0.400005310059
10	0.600001327515	0.399998672485
11	0.599999668121	0.400000331879
12	0.60000008297	0.39999991703
…	0.599999979258	0.400000020742
	0.600000005186	0.399999994814
	0.599999998704	0.400000001296
	0.600000000324	0.399999999676
	0.599999999919	0.400000000081

CT_{t-1}	CT_t	$P(CT_t \mid CT_{t-1})$
C	C	0.5
C	T	0.5
T	C	0.75
T	T	0.25

CT	Col	$P(Col \mid CT_t)$
C	Y	0.1
C	NY	0.9
T	Y	0.75
T	NY	0.25