Graphical Models

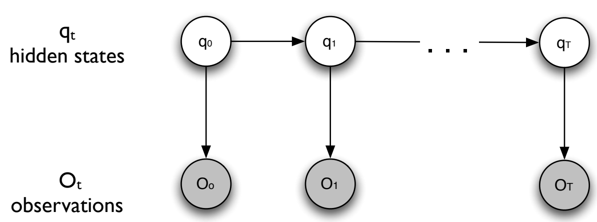
ML 701

Anna Goldenberg

Outline

- Dynamic Models
  - Gaussian Linear Models
    - Kalman Filter
  - DBN
- Undirected Models
- Unification
- Summary

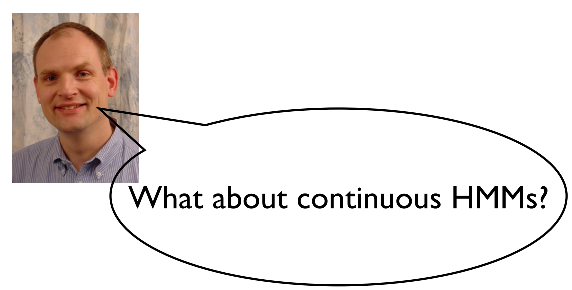
HMMs



$$P(Q,O) = p(q_0) \prod_{t=1}^{T-1} p(q_{t+1}|q_t) \prod_{t=1}^{T} p(O_t|q_t)$$

HMM in short

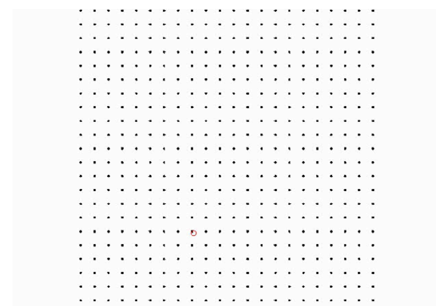
- is a Bayes Net
- satisfies Markov property (independence of states given present)
- with discrete states (time steps are discrete)



Example of use

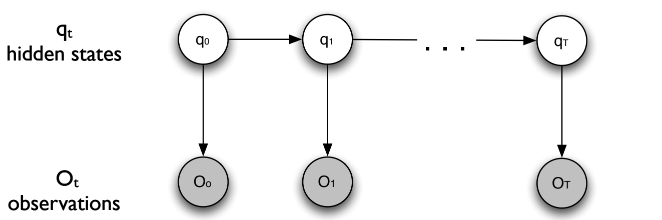
SLAM - Simultaneous Localization and Mapping

http://www.stanford.edu/~paskin/slam/



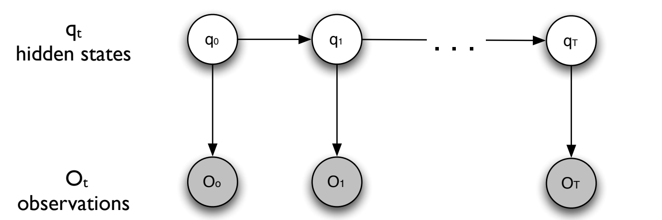
Drawback: Belief State and Time grow quadratically in the number of landmarks

State Space Models



$$P(Q,O) = p(q_0)\prod_{t=1}^{T-1} p(q_{t+1}|q_t)\prod_{t=1}^{T} p(O_t|q_t)$$

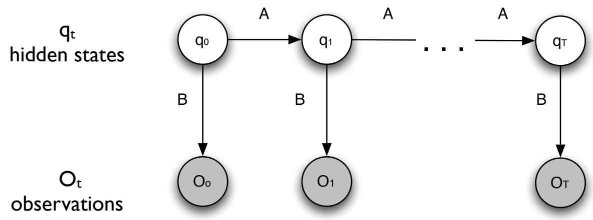
State Space Models



q_t - is a *real-valued* K-dimensional hidden state variable

O_t - is a D-dimensional *real-valued* observation vector

State Space Models



$q_t = f(q_{t-1}) + w_t$ — f determines mean of q_t given mean of q_{t-1}; w_t is zero-mean random noise vector

$O_t = g(q_t) + v_t$ — similarly

Gaussian Linear State Space Models

- O_t and q_t are Gaussian
- f and g are linear and time-invariant

$$q_t = Aq_{t-1} + w_t\,, \qquad w_t \sim N(0, R)$$
$$O_t = Bq_{t-1} + v_t\,, \qquad v_t \sim N(0, S)$$
$$q_0 \sim N(0, \Sigma_0)$$

correction: previously R and S were reversed

A - transition matrix
B - observation matrix

Inference

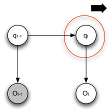
- forward step (filtering)

Kalman Filter

$$p(q_t | O_0, \ldots, O_t)$$

- backward step (smoothing)

$$p(q_t | O_t, O_{t+1}, \ldots, O_T)$$

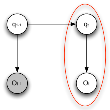


Kalman Filter (1960)

- time update $P(q_{t-1}|O_0, \ldots, O_{t-1}) \to P(q_t|O_0, \ldots, O_{t-1})$

$E(q_{t|t-1}) = A \cdot E(q_{t-1|t-1})$

$V(q_{t|t-1}) = A \cdot V(q_{t-1|t-1})A^T + R$

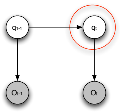


- measurement update $P(q_t|O_o, \ldots, O_{t-1}) \to P(q_t|O_o, \ldots, O_t)$

1. $P(q_t, O_t|O_o, \ldots, O_{t-1})$

$$\begin{bmatrix} E(q_{t|t-1}) \\ B \cdot E(q_{t|t-1}) \end{bmatrix} \quad \begin{bmatrix} V(q_{t|t-1}) & V(q_{t|t-1})B^T \\ BV(q_{t|t-1}) & BV(q_{t|t-1})B^T + R \end{bmatrix}$$

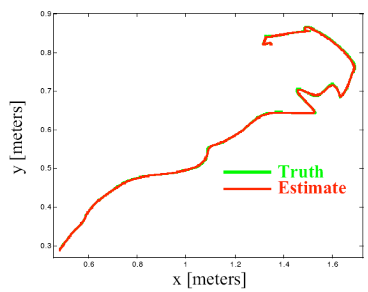
Σ_{11}, Σ_{12}, Σ_{21}, Σ_{22}



2. $P(q_t|O_o, \ldots, O_{t-1}) \to P(q_t|O_o, \ldots, O_t)$

$$E(q_{t|t}) = E(q_{t|t-1}) + \Sigma_{12}\Sigma_{22}^{-1}(O_t - E(O_{t|t}))$$
$$V(q_{t|t}) = V(q_{t|t-1}) - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Example of use



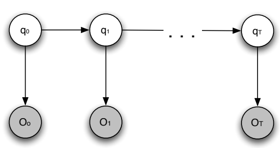
Reported by Welch and Bishop, SIGGRAPH 2001

Kalman Filter Usage

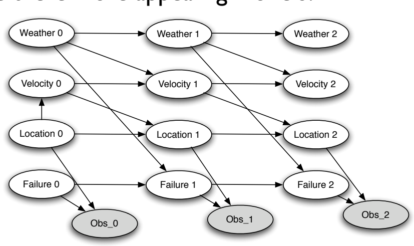
- Tracking motion
  - Missiles
  - Hand motion
  - Lip motion from videos
- Signal Processing
- Navigation
- Economics (for prediction)

Dynamic Bayes Nets

- So far

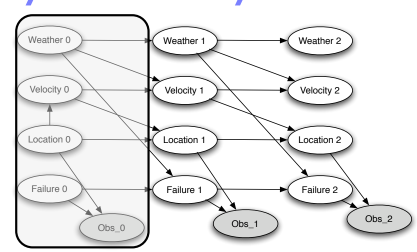


- But are there more appealing models?



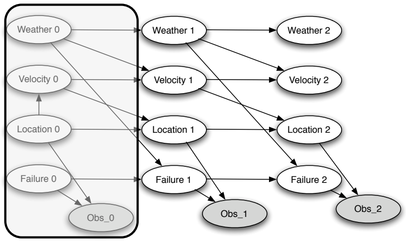
(Koller and Friedman)

Dynamic Bayes Nets



- It's just a Bayes Net!
- Approach to the dynamics
  - 1. Start with some prior for the initial state
  - 2. Predict the next state just using the observation up to the previous time step
  - 3. Incorporate the new observation and re-estimate the current state

Dynamic Bayes Nets



- It's just a Bayes Net!
- Approach to the dynamics
  - 1. Start with some
  - 2. Predict the ne
  - 3. Incorporate the

Most importantly:
Use the structure of the Bayes Net.
Use the independencies!!!

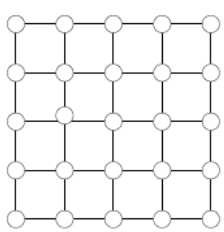
Other graphical models

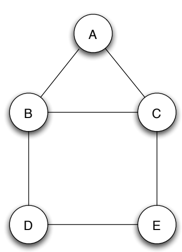
but first...

Any questions so far?

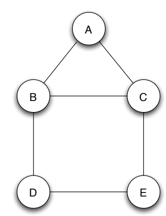
Are all GM directed?

There are Undirected Graphical Models!





Undirected models

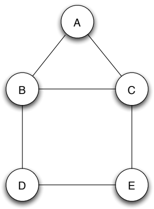


$$p(X) = \frac{1}{Z} \prod_C \psi(X_C)$$

$\psi(X_C)$ - non-negative potential function

What are *C* ?

Cliques



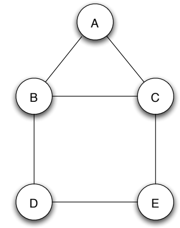
$$p(X) = \frac{1}{Z}\prod_{C}\psi(X_C)$$

$\psi(X_C)$ - non-negative potential function

A clique C is a subset C∈V if ∀i,j∈C, (i,j)∈E

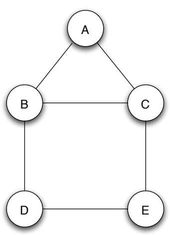
C is maximal if it is not contained in any other clique

Cliques



i) B - a clique?
ii) BC - a maximal clique?
iii) ABCD - a clique?
iv) ABC - a maximal clique?
v) BCDE - a clique?

Decomposition



Note to resolve the confusion:

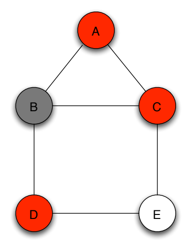
The most common machine learning notation is the decomposition over *maximal* cliques

$$p(A,B,C,D,E) = \frac{1}{Z}p(A,B,C)p(B,D)p(C,E)p(D,E)$$

Independence

Rule: V_1 is independent of V_2 given cutset S
S is called the Markov Blanket (MB)

e.g. MB(B) = {A,C,D}, i.e. the set of neighbors



Are undirected models useful?

- Yes!
  - Used a lot in Physics (Ising model, Boltzmann machine)
  - In vision (every pixel is a node)
  - Bioinformatics

Are undirected models useful?

- Yes!
  - Used a lot in Physics (Ising model, Boltzmann machine)
  - In vision (every pixel is a node), bioinformatics
- Why not more popular?
  - the ZZZZZZ! it's the partition function

$$p(X) = \frac{1}{Z} \prod_{C} \psi(X_C)$$

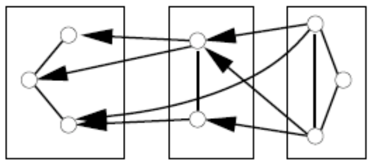
What's Z and ways to fight it

$$Z = \sum_{\forall x} \prod_{C} \psi(X_C)$$

- Approximations
  - Sampling (MCMC sampling is common)
  - Pseudo-Likelihood
  - Mean-field approximation

Chain Graphs

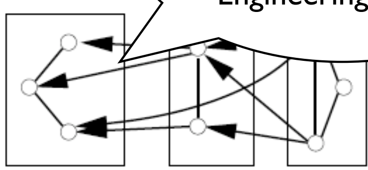
- Generalization of MRFs and Bayes Nets
- Structured as blocks
  - Undirected edges within a block
  - Directed edges between blocks



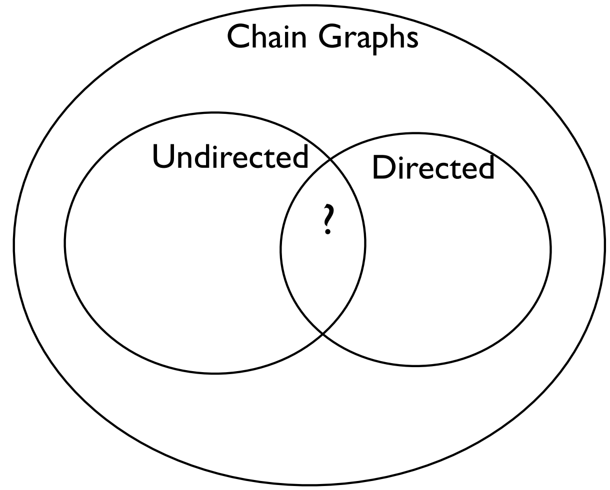
Chain Graphs

- Generalization of MRFs and Bayes Ne
- Structured as blocks
  - Undirected edges withi
  - Directed edges between

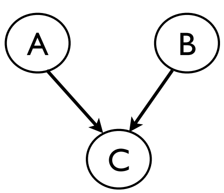
quite intractable
not very popular
used in BioMedical
Engineering (text)



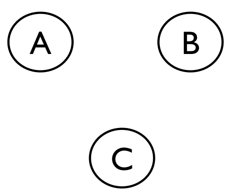
Graphical Models



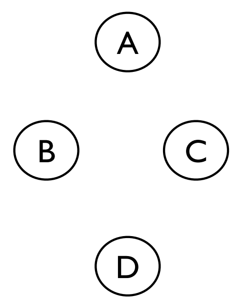
Directed



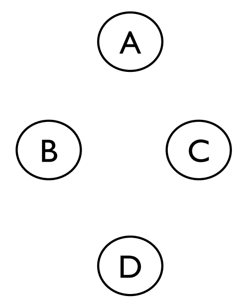
Undirected?

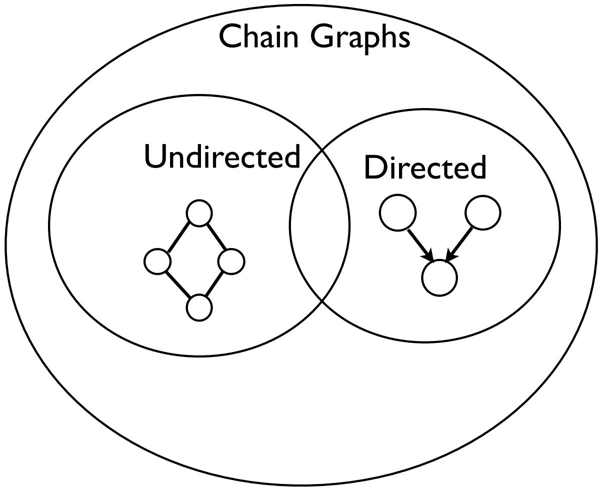


Undirected?



Directed?





Summary

- Graphical Models is a huge evolving field
- There are many other variations that haven't been discussed
- Used extensively in variety of domains
- Tractability issues
- More work to be done!