Markov Models

OSU CS536 Probabilistic Graphical Models

Scott Sanner
Oregon State University
Markov Models (or Markov Chains)

- At each time step, probabilistically transition from current state to next state ($S = \{s_1, s_2, \ldots, s_n\}$)
- Finite State Machine (FSM) view for $n=5$:
Markov Models

• The graphical model view for $t$ steps:

$$P(S_{t+1} | S_t) = P(S_t | S_{t-1})$$

– Note: for $t = \infty$, an infinite graphical model!

• Or assuming transition stationarity, just:
Markov Models

• The Dynamic Bayes Net (DBN) view:
  – State factors into variables: $X_1, X_2, \ldots, X_k$
  – Capture transition independences

\[
P(x_{1}^{t+1}, x_{2}^{t+1}, \ldots, x_{k}^{t+1} \mid x_{1}^{t}, x_{2}^{t}, \ldots, x_{k}^{t}) = \Pi \ldots
\]
Transition Matrix

- Represent $P(s^{t+1}|s^t)$ as transition matrix:

$$T = \begin{bmatrix} P(s_{j}^{t+1}|s_{i}^{t}) 
\end{bmatrix}$$

Do rows or cols $\Sigma=1$?
Transition Probabilities

• Formally
  – Define state set \( S^t = \{s_1, s_2, ..., s_n\} \); \( \forall t \)
  – Define transition matrix \( T_{ij}^t = P(S_i^{t+1}|S_j^t) \); \( \forall t \)

• Properties of \( T_{ij} \)
  – *Stationary*: \( T_{ij}^t = T_{ij}^{t-1} \) OR \( P(S^{t+1}|S^t) = P(S^t|S^{t-1}); \forall t \)
  – *Irreducible*: Possible to get from any \( s_i \) to \( s_j \)
  – *Aperiodic*: Time to return has periodicity = 1
  – *Transient*: Positive probability of not returning to state
  – *Recurrent*: Not transient
  – *Ergodic*: Aperiodic and (positive) recurrent

Examples of each?
Distribution at Time $t$

- Given $P(s^0)$, what is $P(s^t)$?

- Use var. elim. to marginalize over intermediate time steps
  
  
  $P(s^t) = \sum_{s_1, \ldots, s_{t-1}} P(s^0) \prod_{i=0}^{t-1} P(s^{i+1}|s^i)$

- Or let $Ps^0$ & $Ps^t$ be column vectors…
  
  - Then simply: $Ps^t = (T^{^\dagger}) Ps^0$
    
    - Note: Intimate connection between matrix ops and var. elim.
    - When $P(s^{i+1}|s^i)$ factors as a DBN…
      capture many efficiencies of var. elim. via sparse matrix ops

If no evidence after time $t$, all factors for $t+1$ and after marginalize out
Stationary Distribution

- Stationary Distribution $\pi$ at $t=\infty$
  - $\pi = (T^\infty) P s^0$
  - If $T$ ergodic & irreducible, $Ps^0$ irrelevant
    - Reaches *unique* steady-state distribution: $\pi = T \pi$
    - So $\pi=$any row of $T^\infty$
    - Can solve via eigenvector analysis (note: $\lambda=1$)
      - Related to (Krylov) iterated eigenvector computation
    - Or use fixed point to solve linear system
      - $T\pi - \pi = 0 \Rightarrow \pi'T' - \pi'=0 \Rightarrow \pi' (T' - I) = 0$
        s.t. constraints on $\pi$
      - Can solve linear system via matrix inversion

Why? What are they?
Markov Model Applications

• Simple theory, ingenious applications:
  – $n^{th}$-order Markov models
    • Relax Markovian assumption to previous $n$ states
    • Used in text and speech processing
      – N-grams for predicting next word occurrence
        http://nbviewer.ipython.org/gist/yoavg/d76121dfde2618422139
      – Colocation identification
      – Dasher for text input, try it in your web browser
  – More generally
    • Physics (states of systems)
    • Queuing theory (random entries and exits)
    • Economics, Biology, Chemistry, etc…
    • Google!
Google PageRank Example

• Very beautiful use of Markov Models

• Model of web browsing:
  – Probabilistically take link with \( \sim 1/k \) chance if \( k \) links
  – Small chance of random transition

• Stationary distribution \( \pi \) gives PageRank!
  – Measure of “authority”

How to compute on web scale?

Hint: Use iterative method; how would you compute \( T^{256} \)?
Note on Markovian Assumption

• State is only dependent on the preceding $N$ states
  – $N$th-order Markov (often $N=1$)

• Transition probabilities do not change over time: Stationary

• A non-stationary process would be “arrivals at the ground floor elevator in an office building” if time were not in state
  – But augmenting time in the state can make it Markovian
  – In general most processes can be 1st-order Markovian with a properly augmented state
    • How to convert a N-th order Markov Model to 1st-order?

• But state may not always be directly observable...
MCMC

• Markov Chains are also critical for probabilistic inference via MCMC
  – nb. Markov Chain Monte Carlo

• Idea:
  – Given joint over $P(X_1,\ldots,X_n)$
  – Define transition: $T(X_1,\ldots,X_n,X_1',\ldots,X_n')$
    • Which satisfies “detailed balance”
  – Then $\prod (X_1,\ldots,X_n)$ for $T$ is $P(X_1,\ldots,X_n)$!
MCMC Methods

• Three commonly used MCMC techniques:
  – Gibbs sampling
    • Provably satisfies detailed balance
  – Metropolis-Hastings
    • Need to design a symmetric proposal to satisfy detailed balance
  – Hamiltonian Monte Carlo
    • Provably satisfies detailed balance