Applied Machine Learning

Markov Chains II

Markov Chains II

- Simulation
- Stationary distribution
- M-grams

Markov Chains: States and Transition Probabilities

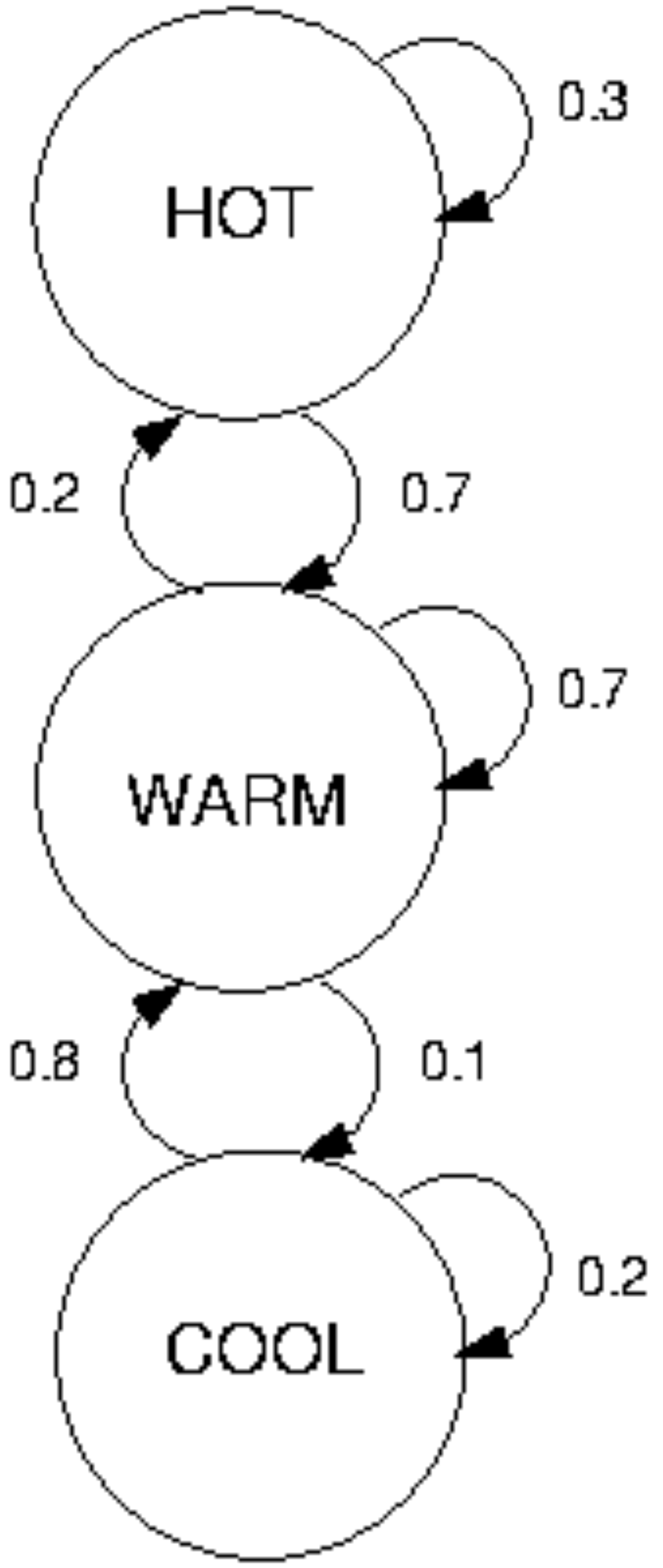
- State encoding:
  [hot $= 0$, warm $= 1$, cool $= 2$]

- Transition Probability Matrix P

  - $p_{i,j} = P(X_n = j \mid X_{n-1} = i)$

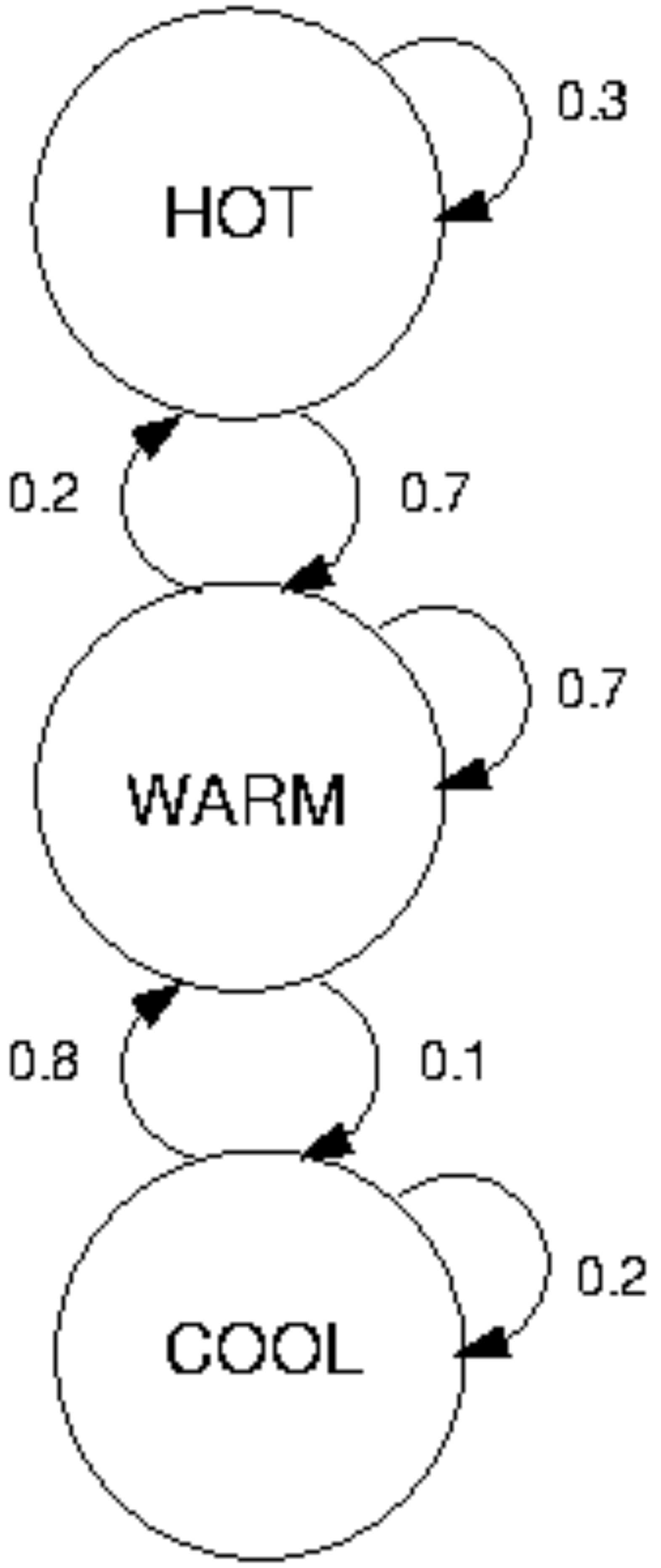
  - $P = \begin{bmatrix} 0.3 & 0.7 & 0 \\ 0.2 & 0.7 & 0.1 \\ 0 & 0.8 & 0.2 \end{bmatrix}$

  - $\sum_j P(x_n = j \mid X_{n-1} = i) = \sum_j p_{i,j} = 1$



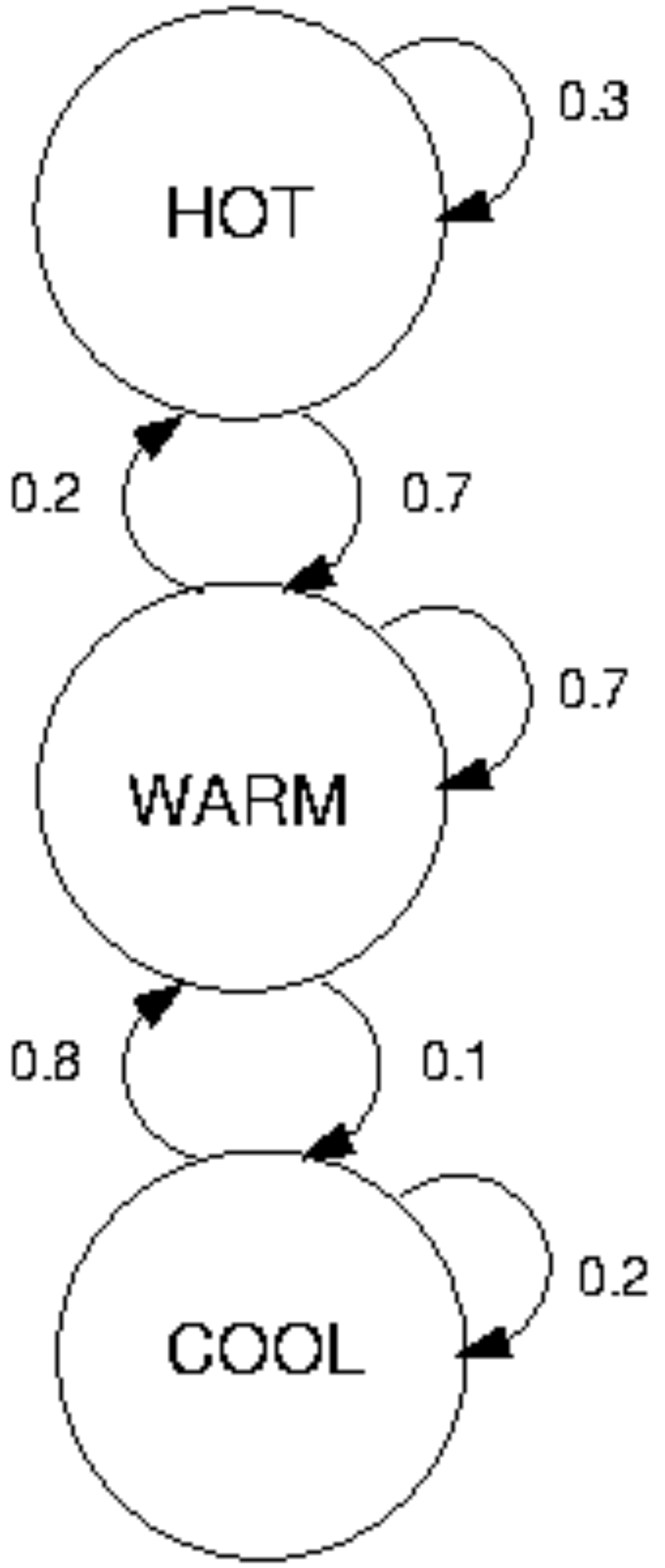
Simulating Probability Distributions

- Initial probability distribution $P(X_0 = i)$
    - $\pi_{[1 \times k]}$
    - $\pi = [0.0, 0.20, 0.80]$
- Probability distribution at time 1:
    - $$\begin{aligned} P(X_1 = j) &= \sum_i P(X_1 = j, X_0 = i) \\ &= \sum_i P(X_1 = j \mid X_0 = i) P(X_0 = 1) \\ &= \sum_i p_{i,j} \pi_i \end{aligned}$$
    - As matrix multiplication: $\pi_1 = \pi P$
    - $\pi_1 = [0.04, 0.78, 0.18]$

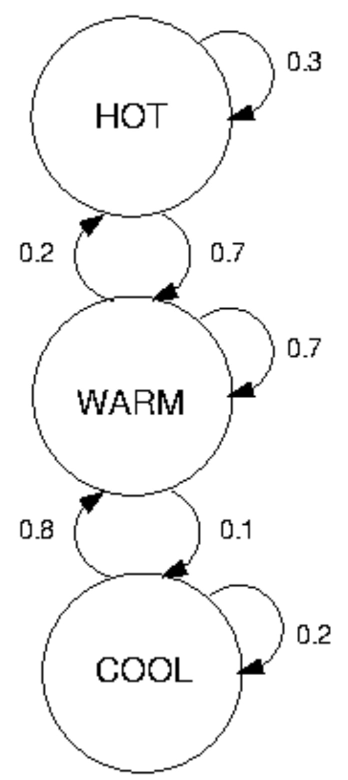


Simulating Probability Distributions

- Initial probability distribution $P(X_0 = i)$
  - $\pi_{[1 \times k]}$
  - $\pi = [0.0, 0.20, 0.80]$
- Probability distribution at time 2:
  - $$\begin{aligned} P(X_2 = j) &= \textstyle\sum_i P(X_2 = j, X_1 = i) \\ &= \textstyle\sum_i P(X_2 = j | X_1 = i) P(X_1 = 1) \\ &= \textstyle\sum_i p_{i,j} (\sum_{k,i} p_{k,j} \pi_k) \end{aligned}$$
  - As matrix multiplication: $\pi_n = \pi_{n-1} P = \pi P^n$
  - $\pi_1 = [0.04, 0.78, 0.18]$, $\pi_{10} = [0.20, 0.71, 0.09]$



Stationary Distribution



- After a number of time steps
  - stuck at absorbing state
  - stabilize at some distribution
- Irreducible chain
  - finite number of states
  - no isolated state, so no absorbing state and all the states have probability greater than 0
  - evolve into stationary distribution s
    - $\lim_{n\to\infty} \pi P^n = s$

- Probability of states after enough steps to make it independent on where the walk started
  - Probability of reaching each state
- Advancing one more step leads to the same distribution s
  - $sP = s$
    - s is eigenvector of $P^\top$ with eigenvalue 1 divided by 1-norm, so probabilities sum 1.
    - if irreducible, there is only one such eigenvector
- $s^\top = [0.20, 0.71, 0.09]$

M-Order Markov Models

- Sentence: “Markov chains are cool”
- Markov chain to produce text
- Order 0: Single elements, no dependency
- Order 1: Dependency from previous element
  - pairs
  - bi-grams
- Order 2: Dependence from two previous elements
  - triplets
  - tri-grams
- Order M: M-gram

M-Grams

- Produce sequences of bits, characters, words
- Identify likely errors in sequences of bits, characters, words
- Text examples
  - frequency of M-grams, probability of transitions
  - there may be rare M-grams, even unseen M-grams
- Smoothing
  1. fixed low probability for each unseen M-gram
  2. distribute fixed probability among all unseen M-grams
     - Issue: frequency of unseen M-grams vary

Markov Chains II

- Simulation
- Stationary distribution
- M-grams

Applied Machine Learning

Markov Chains II