# Sequential Data and Markov Models

# Sargur N. Srihari

srihari@cedar.buffalo.edu

# Machine Learning Course:

http://www.cedar.buffalo.edu/~srihari/CSE574/index.html

# Sequential Data Examples

- Often arise through measurement of time series
  - -Snowfall measurements on successive days
  - -Rainfall measurements on successive days
  - Daily values of currency exchange rate
  - Acoustic features at successive time frames in speech recognition
  - Nucleotide base pairs in a strand of DNA
  - -Sequence of characters in an English sentence
  - Parts of speech of successive words

### Markov Model – Weather

• The weather of a day is observed as being one of the following:

—State 1: Rainy

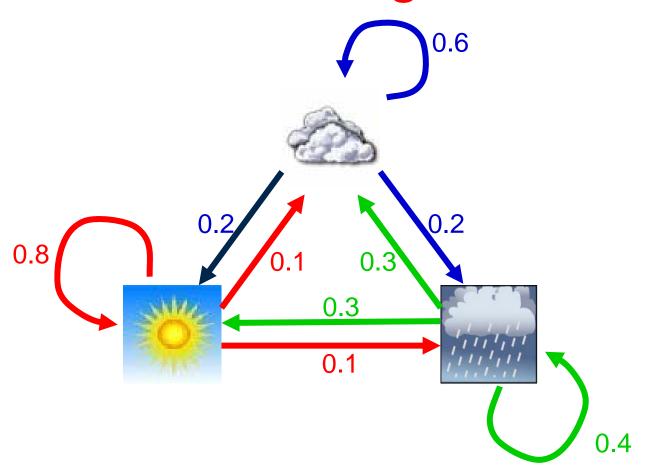
—State 2: Cloudy

-State 3: Sunny

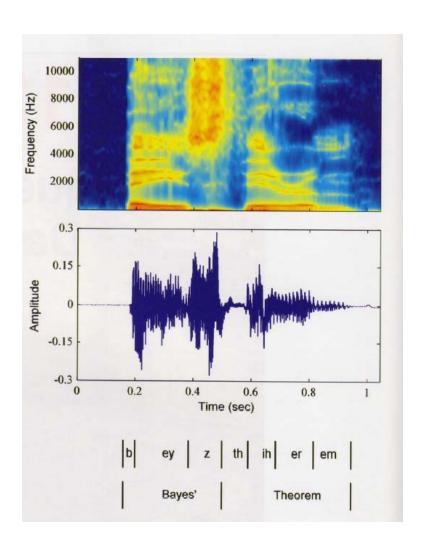


|       |        | Tomorrow |        |       |
|-------|--------|----------|--------|-------|
|       |        | Rain     | Cloudy | Sunny |
| Today | Rain   | 0.3      | 0.3    | 0.4   |
|       | Cloudy | 0.2      | 0.6    | 0.2   |
|       | Sunny  | 0.1      | 0.1    | 0.8   |

# Markov Model – Weather State Diagram



# Sound Spectrogram of Speech

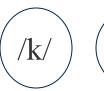


- "Bayes Theorem"
- Plot of the intensity of the spectral coefficients versus time index
- Successive observations of speech spectrum highly correlated (Markov dependency)

# Markov model for the production of spoken words

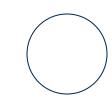
- States represent phonemes
- Production of word: "cat"
- Represented by states /k/ /a/ /t/
- Transitions from
  - /k/ to /a/
  - /a/ to /t/
  - /t/ to a silent state
- Although only correct cat sound is represented by model, perhaps other transitions can be introduced,
  - eg, /k/ followed by /t/

Markov Model for word "cat"









## Stationary vs Non-stationary

- Stationary: Data evolves over time but distribution remains same
  - —e.g., dependence of current word over previous word remains constant
- Non-stationary: Generative distribution itself changes over time

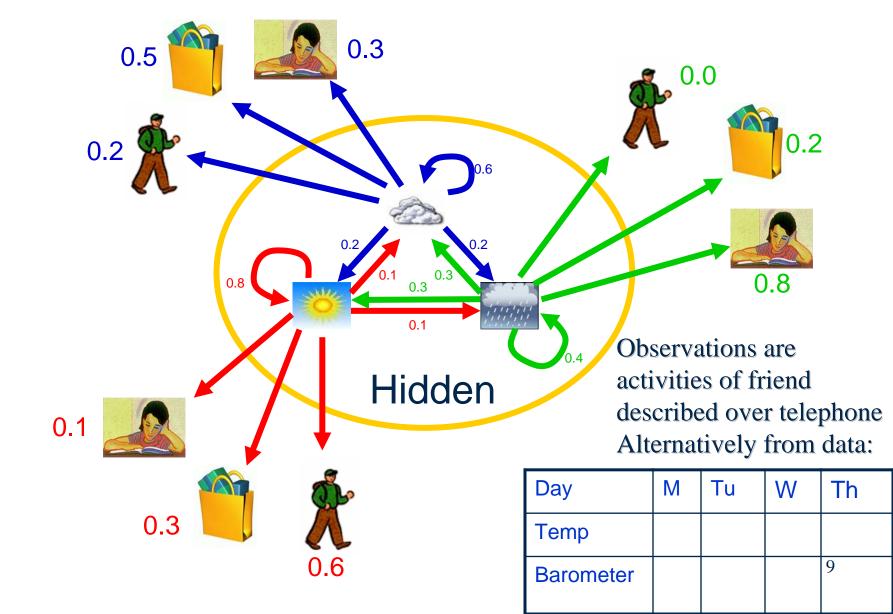
# Making a Sequence of Decisions

- Processes in time, states at time t are influenced by a state at time t-1
- Wish to predict next value from previous values, e.g., financial forecasting
- Impractical to consider general dependence of future dependence on all previous observations
  - Complexity grows without limit as number of observations increases
- Markov models assume dependence on most recent observations

#### Latent Variables

- While Markov models are tractable they are severely limited
- Introduction of latent variables provides a more general framework
- Lead to state-space models
- When latent variables are:
  - —Discrete
    - they are called *Hidden Markov models*
  - —Continuous
    - they are linear dynamical systems

## Hidden Markov Model



## Markov Model Assuming Independence









- —Assume observations are independent
- Graph without links
- To predict whether it rains tomorrow is only based on relative frequency of rainy days
- Ignores influence of whether it rained the previous day

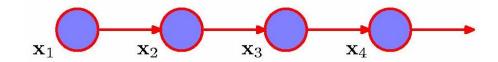
#### Markov Model

- Most general Markov model for observations  $\{x_n\}$
- Product rule to express joint distribution of sequence of observations

$$p(x_1,...x_N) = \prod_{n=1}^{N} p(x_n \mid x_1,...x_{n-1})$$

#### First Order Markov Model

• Chain of observations  $\{x_n\}$ 



Joint distribution for a sequence of n variables

$$p(x_1,...x_N) = p(x_1) \prod_{n=2}^{N} p(x_n | x_{n-1})$$

It can be verified (using product rule from above) that

$$p(x_n \mid x_1..x_{n-1}) = p(x_n \mid x_{n-1})$$

- If model is used to predict next observation, distribution of prediction will only depend on preceding observation and independent of earlier observations
- Stationarity implies conditional distributions  $p(x_n/x_{n-1})$  are all equal

# Markov Model – Sequence probability

 What is the probability that the weather for the next 7 days will be "S-S-R-R-S-C-S"?

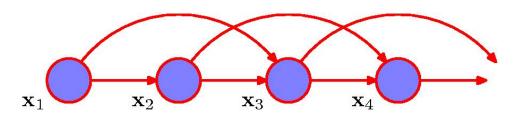
$$O = \{S_3, S_3, S_3, S_1, S_1, S_1, S_3, S_2, S_3\}$$

—Find the probability of *O*, given the model.

$$\begin{split} P(O \mid Model \,) &= P(S_3, S_3, S_3, S_1, S_1, S_3, S_2, S_3 \mid Model \,) \\ &= P(S_3) \cdot P(S_3 \mid S_3) \cdot P(S_3 \mid S_3) \cdot P(S_1 \mid S_3) \\ &\cdot P(S_1 \mid S_1) \cdot P(S_3 \mid S_1) \cdot P(S_2 \mid S_3) \cdot P(S_3 \mid S_2) \\ &= \pi_3 \cdot a_{33} \cdot a_{33} \cdot a_{31} \cdot a_{11} \cdot a_{13} \cdot a_{32} \cdot a_{23} \\ &= 1 \cdot (0.8) \cdot (0.8) \cdot (0.1) \cdot (0.4) \cdot (0.3) \cdot (0.1) \cdot (0.2) \\ &= 1.536 \times 10^{-4} \end{split}$$

### Second Order Markov Model

• Conditional distribution of observation  $x_n$  depends on the values of two previous observations  $x_{n-1}$  and  $x_{n-2}$ 



$$p(x_1,..x_N) = p(x_1)p(x_2 \mid x_1) \prod_{n=3}^{N} p(x_n \mid x_{n-1}, x_{n-2})$$

Each observation is influenced by previous two observations

### M<sup>th</sup> Order Markov Source

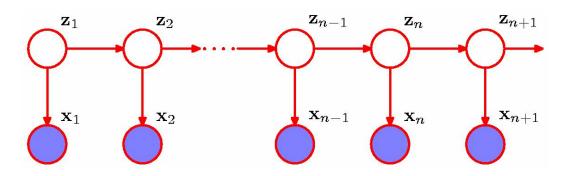
- Conditional distribution for a particular variable depends on previous M variables
- Pay a price for number of parameters
- Discrete variable with K states
  - First order:  $p(x_n/x_{n-1})$  needs K-1 parameters for each value of  $x_{n-1}$  for each of K states of  $x_n$  giving K(K-1) parameters
  - $-M^{th}$  order will need  $K^{M-1}(K-1)$  parameters

# Introducing Latent Variables

- Model for sequences not limited by Markov assumption of any order but with limited number of parameters
- For each observation  $x_n$ , introduce a latent variable  $z_n$
- $z_n$  may be of different type or dimensionality to the observed variable
- Latent variables form the Markov chain
- Gives the "state-space model"

Latent variables

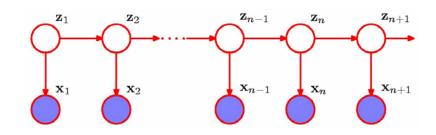
Observations



#### Conditional Independence with Latent Variables

 Satisfies key assumption that

$$Z_{n+1} \perp Z_{n-1} \mid Z_n$$



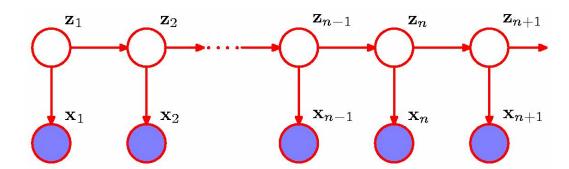
From d-separation

When latent node  $z_n$  is filled, the only path between  $z_{n-1}$  and  $z_{n+1}$  has a head-to-tail node that is blocked

#### Jt Distribution with Latent Variables

Latent variables

**Observations** 



Joint distribution for this model

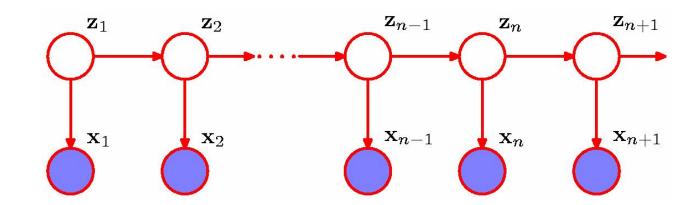
$$p(x_1,...x_N,z_1,...z_n) = p(z_1) \left[ \prod_{n=2}^N p(z_n \mid z_{n-1}) \right] \prod_{n=1}^N p(x_n \mid z_n)$$

- There is always a path between any  $x_n$  and  $x_m$  via latent variables which is never blocked
- Thus predictive distribution  $p(x_{n+1}|x_1,...,x_n)$  for observation  $x_{n+1}$  does not exhibit conditional independence properties and is hence dependent on all previous observations

## Two Models Described by Graph

Latent variables

**Observations** 



- Hidden Markov Model: If latent variables are discrete:
  Observed variables in a HMM may be discrete or continuous
- Linear Dynamical Systems: If both latent and observed variables are Gaussian

# Further Topics on Sequential Data

Hidden Markov Models:

http://www.cedar.buffalo.edu/~srihari/CSE574/Chap11/Ch11.2-HiddenMarkovModels.pdf

Extensions of HMMs:

http://www.cedar.buffalo.edu/~srihari/CSE574/Chap11/Ch11.3-HMMExtensions.pdf

Linear Dynamical Systems:

http://www.cedar.buffalo.edu/~srihari/CSE574/Chap11/Ch11.4-LinearDynamicalSystems.pdf

Conditional Random Fields:

http://www.cedar.buffalo.edu/~srihari/CSE574/Chap11/Ch11.5-ConditionalRandomFields.pdf