EE211: Robotic Perception and Intelligence

Lecture 12 Markov Models

Jiankun WANG

Department of Electronic and Electrical Engineering
Southern University of Science and Technology

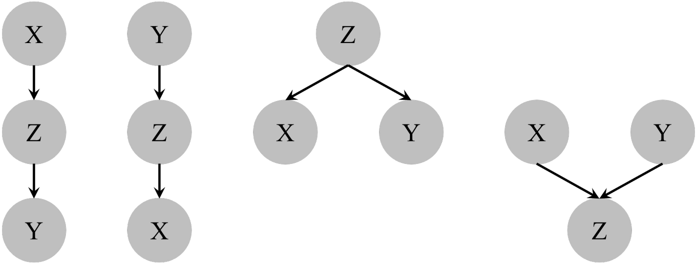
Undergraduate Course, Dec 2024

Outline

Outline

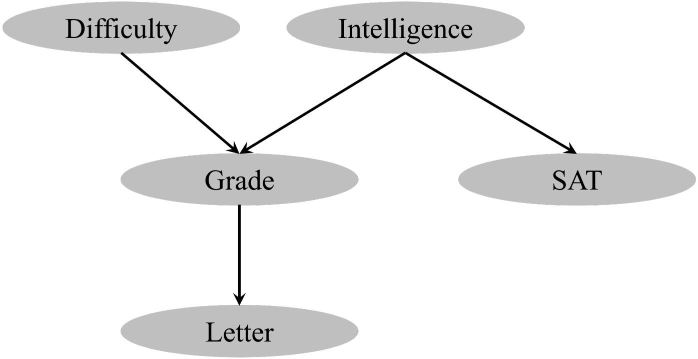
Indirect Connection between Two Variables

- Three (Four) Cases
  - Indirect causal/evidential effect
  - Common cause
  - Common effect

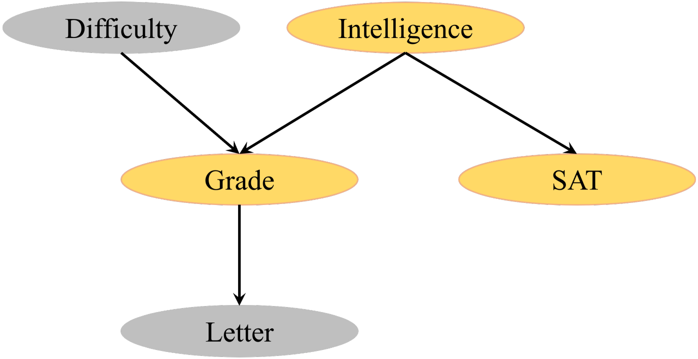


https://cg.csc.lic.ac.uk/ xiaowei/.

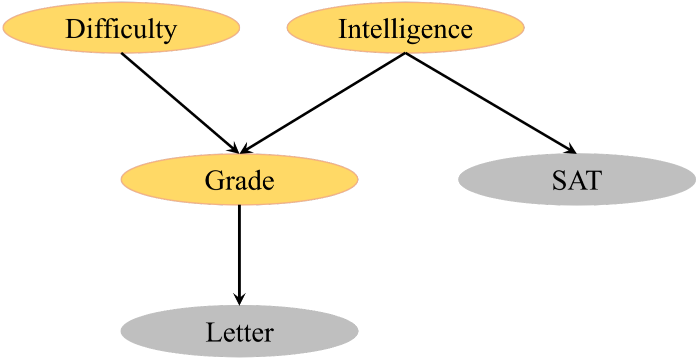
Example



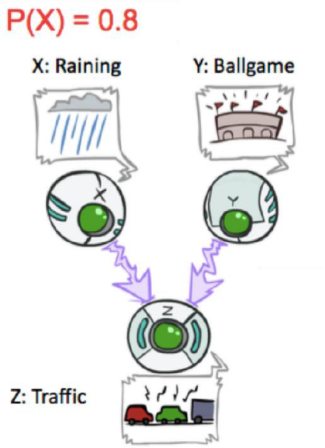
Example: Common Cause



Example: Common Effect

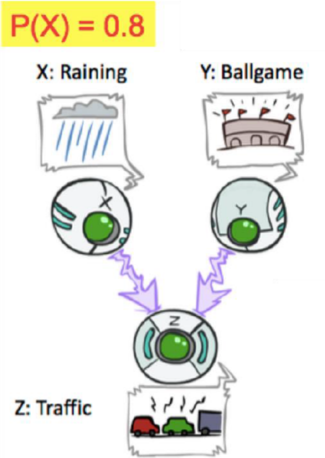


Example: Common Effect - 1



X	Y	Z	P
T	T	T	0.076
T	T	F	0.004
T	F	T	0.576
T	F	F	0.144
F	T	T	0.162
F	T	F	0.002
F	F	T	0.090
F	F	F	0.009

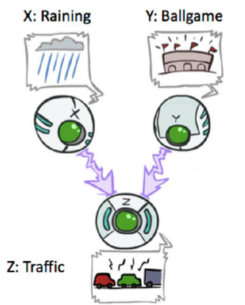
Example: Common Effect - 2



X	Y	Z	P
T	T	T	0.076
T	T	F	0.004
T	F	T	0.576
T	F	F	0.144
F	T	T	0.162
F	T	F	0.002
F	F	T	0.090
F	F	F	0.009

Example: Common Effect - 3

P(X) = 0.8



X	Y	Z	P
T	T	T	0.076
T	T	F	0.004
T	F	T	0.576
T	F	F	0.144
F	T	T	0.018
F	T	F	0.002
F	F	T	0.090
F	F	F	0.009

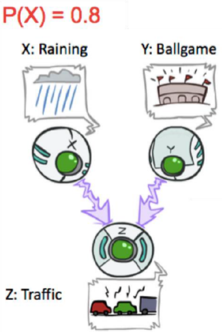
$$P(X|Y) = \frac{0.076+0.004}{0.076+0.004+0.018+0.002}$$

$$= 0.08 / 0.1$$

$$= 0.8$$

X and Y are independent!

But Suppose Also Know Z=T



$$P(X|Z) = \frac{.076+.576}{.076+.576+.018+.090} = 0.652/0.76 = 0.858$$

$$P(X|Y,Z) = \frac{0.076}{0.076+0.018} = 0.8085$$

X	Y	Z	P
T	T	T	0.076
T	T	F	0.004
T	F	T	0.576
T	F	F	0.144
F	T	T	0.018
F	T	F	0.002
F	F	T	0.090
F	F	F	0.009

X and Y are not independent given Z!

Summary

- Causal trail: $X \to Z \to Y$: active iff Z not observed
- Evidential Trail: $X \leftarrow Z \leftarrow Y$: active iff Z is not observed
- Common Cause: $X \leftarrow Z \to Y$: active iff Z is not observed
- Common Effect: $X \to Z \leftarrow Y$: active iff either Z or one of its descendants is observed

Outline

- Joint distribution for a sequence of observations:

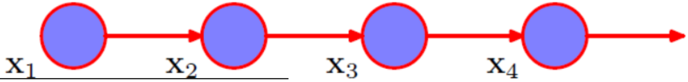
$$p(x_1, ..., x_N) = \prod_{n=1}^{N} p(x_n|x_1, ..., x_{n-1}).$$

- Each of the conditional distributions is independent of all previous observations except the most recent, we get first order Markov chain:

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

- So the conditional distribution for observation x_n, given all of the observations up to time n:

$$p(x_n|x_1, ..., x_{n-1}) = p(x_n|x_{n-1}).$$



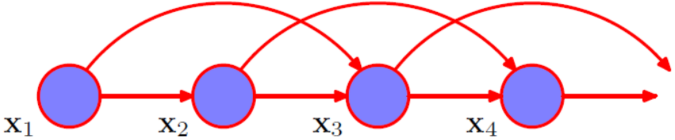
Pattern Recognition and Machine Learning.

Second-order Markov Chain

- For many sequential observations, we anticipate that the trends in the data over several successive observations will provide important information in predicting the next value.
- If we allow the predictions to depend also on the previous-but-one value, we obtain a second-order Markov chain:

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

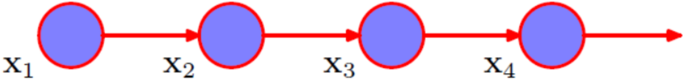
where the conditional distribution of x_n given x_{n-1} and x_{n-2} is independent of all observations $x_1, ..., x_{n-3}$.



Exercise 1

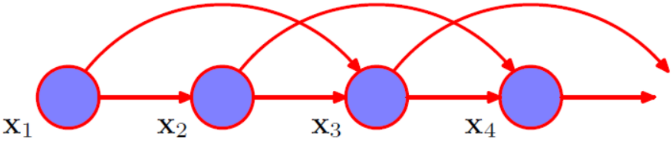
- Verify that the Markov model in following figure satisfies the conditional independence properties

$$p(x_n|x_1, \ldots, x_{n-1}) = p(x_n|x_{n-1}).$$



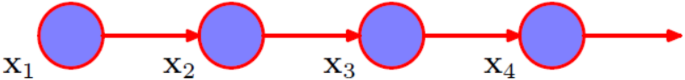
- Verify that the Markov model in following figure satisfies the conditional independence properties

$$p(x_n|x_1, \ldots, x_{n-1}) = p(x_n|x_{n-1}, x_{n-2}).$$



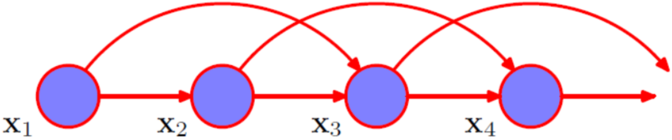
Solution 1

- Since the arrows on the path from x_m to x_n, with $m < n-1$, will meet head-to-tail at x_{n-1}, which is in the conditioning set, all such paths are blocked by x_{n-1} and hence $p(x_n|x_1, ..., x_{n-1}) = p(x_n|x_{n-1})$ holds.



- The same argument applies in the case depicted in following figure, with the modification that $m < n-2$ and that paths are blocked by x_{n-1} or x_{n-2}, so the following equation holds

$$p(x_n|x_1, ..., x_{n-1}) = p(x_n|x_{n-1}, x_{n-2}).$$



Discussion

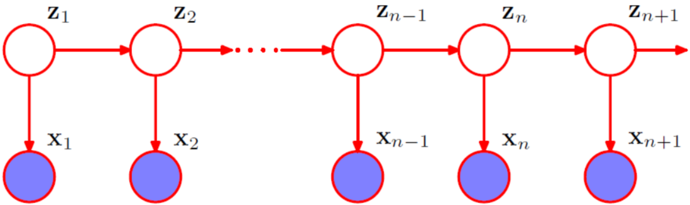
- For an M^{th} order Markov chain, the conditional distribution for a particular variable depends on the previous M variables.
- Suppose the observations are discrete variables having K states. Then the conditional distribution $p(x_n|x_{n-1})$ in a first-order Markov chain will be specified by a set of $K-1$ parameters for each of the K states of x_{n-1} giving a total of $K(K-1)$ parameters.
- An M^{th} order Markov chain will need $K^{M-1}(K-1)$ parameters.

Outline

Overview

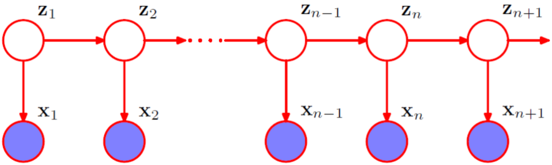
- For each observation x_n, we introduce a corresponding latent variable z_n, and it satisfies the key conditional independence property that z_{n-1} and z_{n+1} are independent given z_n, so that $z_{n+1} \perp\!\!\!\perp z_{n-1} | z_n$.
- The joint distribution for this model is

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



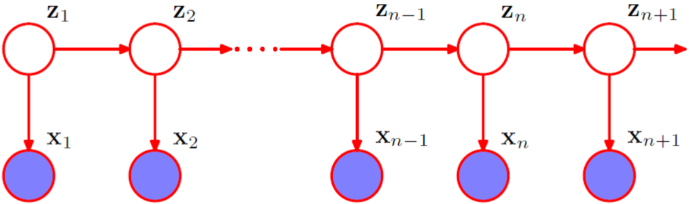
Hidden Markov Models

- Using the d-separation criterion, we see that there is always a path connecting any two observed variables x_n and x_m via the latent variables, and that this path is never blocked.
- Thus the predictive distribution $p(x_{n+1}|x_1, ..., x_n)$ for observation x_{n+1} given all previous observations does not exhibit any conditional independence properties, and so our predictions for x_{n+1} depends on all previous observations.
- The observed variables, however, do not satisfy the Markov property at any order.



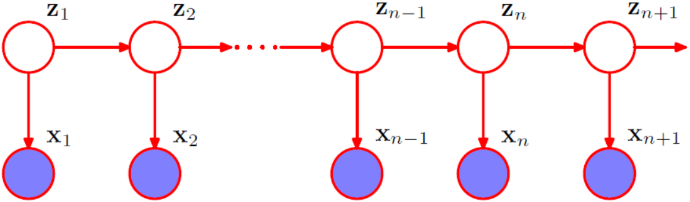
Exercise 2

- By using d-separation, show that the distribution $p(x_1, \ldots, x_N)$ of the observed data for the state space model represented by the directed graph in following Figure does not satisfy any conditional independence properties and hence does not exhibit the Markov property at any finite order.



Solution 2

- From following figure we see that for any two variables x_n and $x_m, m \neq n$, there is a path between the corresponding nodes that will only pass through one or more nodes corresponding to z variables. None of these nodes will be in the conditioning set and the edges on the path meet head-to-tail. Thus, there will be an unblocked path between x_n and x_m and the model will not satisfy any conditional independence or finite order Markov properties.



Transition Probabilities

- $p(z_n|z_{n-1})$: probability distribution of z_n depends on the state of the previous latent variable z_{n-1}.
- The latent variables are K-dimensional binary variables, this conditional distribution corresponds to a table of numbers that we denote by A, the elements of which are known as transition probabilities:

$$A_{jk} = p(z_{nk} = 1|z_{n-1,j} = 1),\ 0 \leq A_{jk} \leq 1, \sum_k A_{jk} = 1,$$

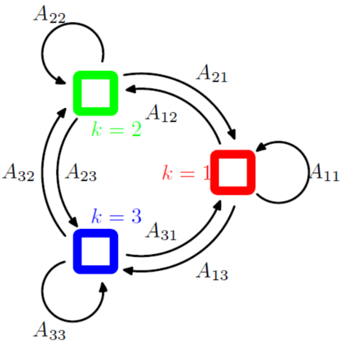
so that A has $K(K-1)$ independent parameters. And we get

$$p(z_n|z_{n-1}, A) = \prod_{k=1}^{K} \prod_{j=1}^{K} A_{jk}^{z_{n-1,j} z_{nk}}.$$

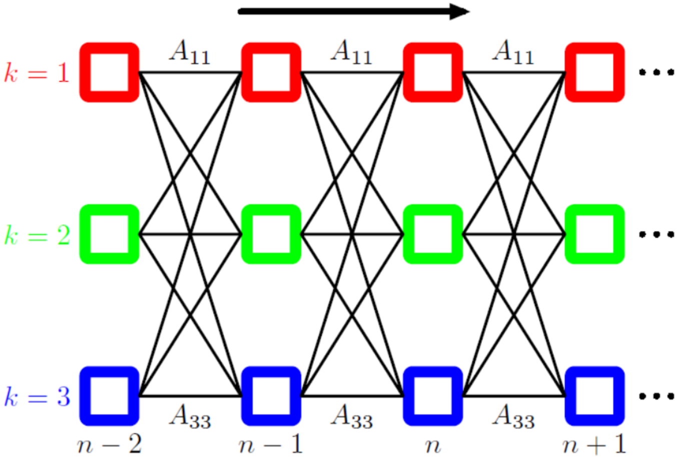
Transition Probabilities

- The initial latent node z_1 is special in that it does not have a parent node, and so it has a marginal distribution $p(z_1)$ represented by a vector of probabilities π with elements $\pi_k = p(z_{1k} = 1)$, so that

$$p(z_1|\pi) = \prod_{k=1}^{K} \pi_k^{z_{1k}}, \sum_k \pi_k = 1.$$



Transition Probabilities



Emission Probabilities

- $p(x_n|z_n, \phi)$: conditional distributions of the observed variables, ϕ is a set of parameters governing the distribution.
- Because x_n is observed, the distribution $p(x_n|z_n, \phi)$ consists, for a given value of ϕ, of a vector of K numbers corresponding to the K possible states of the binary vector z_n.
- So we represent the emission probabilities:

$$p(x_n|z_n, \phi) = \prod_{k=1}^{K} p(x_n|\phi_k)^{z_{nk}}.$$

Joint Probability Distribution

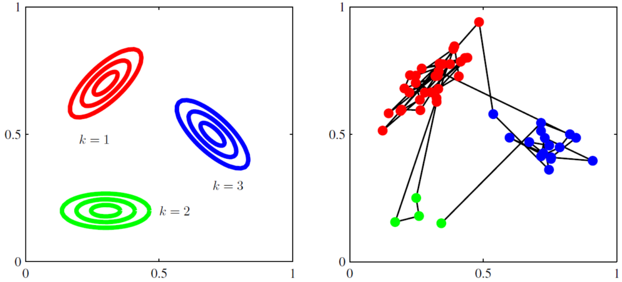
- The joint probability distribution over both latent and observed variables is given by

$$p(X, Z|\phi) = p(z_1|\pi)\left[\prod_{n=2}^{N} p(z_n|z_{n-1}, A)\right]\prod_{m=1}^{N} p(x_m|z_m, \phi),$$

where $X = \{x_1, ..., x_N\},\ Z = \{z_1, ..., z_N\},\ \theta = \{\pi, A, \phi\}$ denotes the set of parameters governing the model.

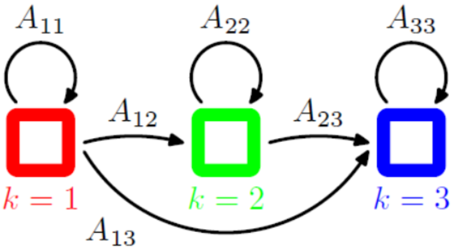
Better Understanding

- Illustration of sampling from a hidden Markov model having a 3-state latent variable z and a Gaussian emission model $p(x|z)$ where x is 2-dimensional.
- Transition matrix is fixed so that in any state there is a 5% probability of making a transition to each of the other states, and consequently a 90% probability of remaining in the same state.



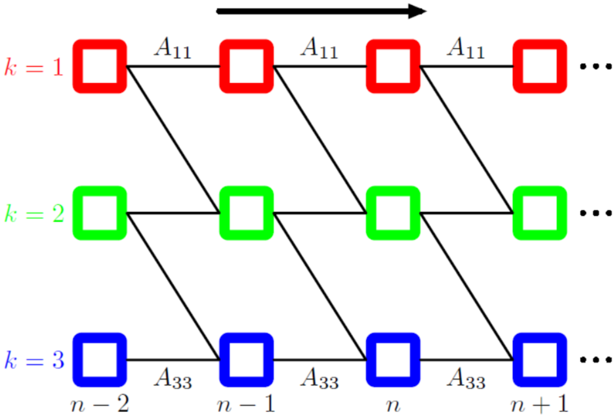
Left-to-Right HMM

- Set the elements A_{jk} of A to zero if $k < j$.
- Initial state probabilities for $p(z_1)$ are modified so that $p(z_{11}) = 1$, $p(z_{1j}) = 0$ for $j \neq 1$, in other words every sequence is constrained to start in state $j = 1$.



Left-to-Right HMM

- Further constraints: large changes in the state index do not occur, so that $A_{jk} = 0$ if $k > j + \Delta$.



Exercise

Define an HMM model with three latent states $\{A, B, C\}$ and observations $\{0, 1, 2\}$. The initial stable probabilities are $\pi_A = 1$ and $\pi_B = \pi_C = 0$. The transition and emission probabilities are as follows:

	A	B	C	0	1	2
A	0.2	0.8	0.0	0.8	0.2	0.0
B	0.0	0.8	0.2	0.0	0.6	0.4
C	0.4	0.0	0.6	0.2	0.0	0.8

(1) Draw the state diagram of this HMM and show the transition probabilities. (2) Give all state paths with non-zero probability for the sequence $O = 0, 1, 2$.

Q & A

1. Preliminaries: d-separation criterion
2. Markov Models
3. Hidden Markov Models