Probabilistic Graph Models: Factor Graphs

ECE/CS 498 DS U/G

Lecture 16

Ravi K. Iyer

Dept. of Electrical and Computer Engineering

University of Illinois at Urbana Champaign

Announcements

- MP2 Checkpoint 3 is due tonight
- MP3 to be released on Friday, Mar 29
- Discussion Section on Friday, Mar 29
  - A problem on Factor Graphs

Hidden Markov Models

Model

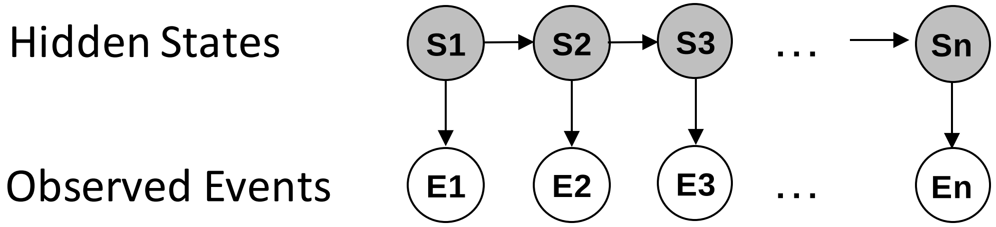
- Set of hidden states $\boldsymbol{S} = \{\boldsymbol{\sigma_1}, \dots, \boldsymbol{\sigma_N}\}$
- Set of observable events $\boldsymbol{E} = \{\boldsymbol{\epsilon_1}, \dots, \boldsymbol{\epsilon_M}\}$
- Transition probability matrix $\boldsymbol{A}$
- Observation matrix $\boldsymbol{B}$
- Initial distribution of hidden states $\boldsymbol{\pi}$

Model assumptions

- An observation depends on its hidden state
- A state variable only depends on the immediate previous state (Markov assumption)
- The future observations and the past observations are conditionally independent given the current hidden state

Advantages:

- HMM can model sequential nature of input data (future depends on the past)
- HMM has a linear-chain structure that clearly separates system state and observed events.

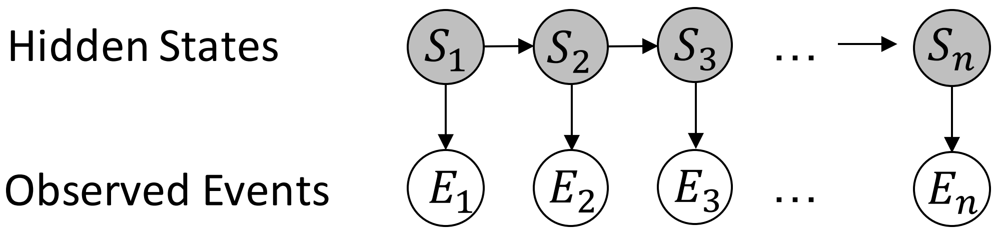


A Hidden Markov model on observed events and system states

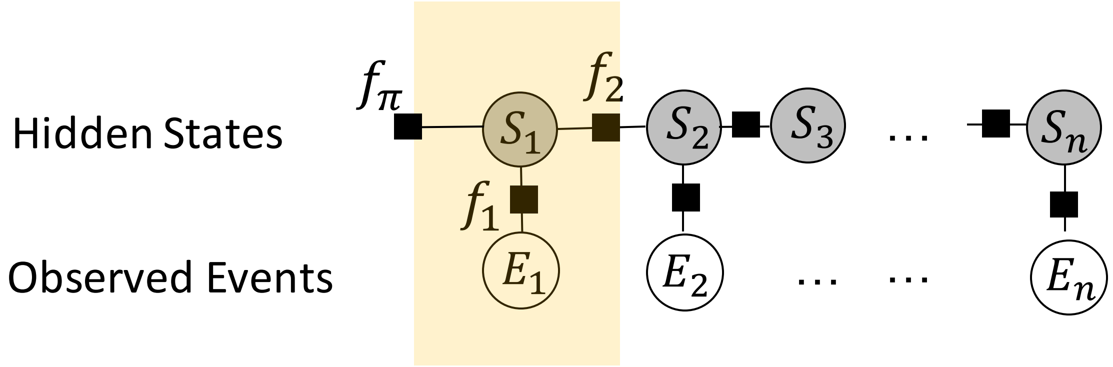
$$P(S_1, \dots, S_n, E_1, \dots, E_n)$$
$$= P(S_1)P(E_1|S_1)\prod_{i=2}^{n} P(S_i|S_{i-1})P(E_i|S_i)$$

Conversion of a Hidden Markov Model to a Factor Graph

Hidden Markov Model

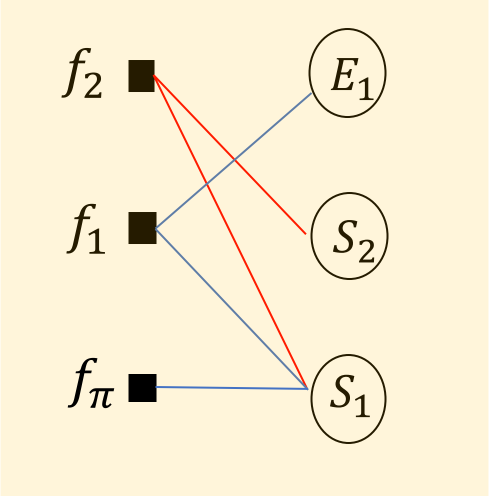


Factor Graph of the HMM



The above *Factor Graph* (FG) is a generalization of the Hidden Markov Model

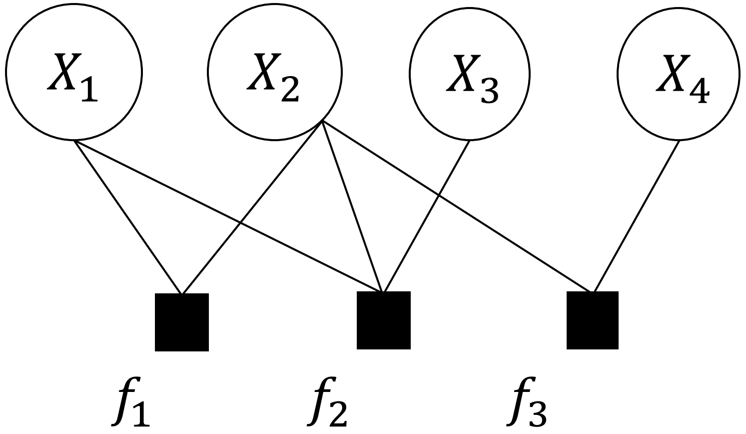
- Boxes (f_π, f_1, f_2) represents factor function
- In the above case, it maintains the Markov assumption between states



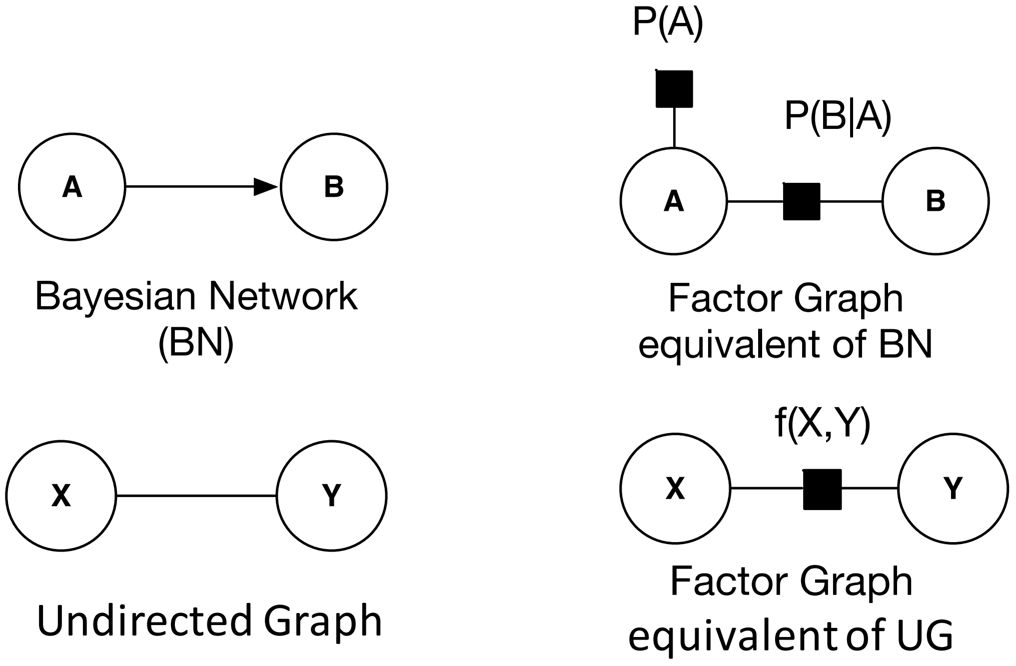
Bipartite graph representation of the FG

Definition of a Factor Graph

A factor graph is a **bipartite, undirected graph** of **random variables and factor functions. [Frey et. al. 01]**



*FG can represent both **causal and non-causal relations**.*



Example Factor function for HMMs

Assume that the state space and observation space are $S = \{\sigma_0, \sigma_1\}$, $E = \{\epsilon_1, \epsilon_2\}$. An example of factor functions is shown.

S	$f_\pi(S)$
σ_0	40
σ_1	25

S_t	E_t	$f_1(S_t, E_t)$
σ_0	ϵ_1	20
σ_0	ϵ_2	15
σ_1	ϵ_1	40
σ_1	ϵ_2	3

S_t	S_{t+1}	$f_2(S_t, S_{t+1})$
σ_0	σ_0	5
σ_0	σ_1	1
σ_1	σ_0	10
σ_1	σ_1	15

- Factor values represents the *affinities* between the related variables
  - E.g., $f_1(\sigma_1, \epsilon_1) > f_1(\sigma_0, \epsilon_1)$ implies that σ_1 and ϵ_1 are more compatible than σ_0 and ϵ_1
- Factor functions don't necessarily represent CPDs or joint probability distributions
- How are these values found?
  1. Given by expert or from domain knowledge
  2. Derived from the data (priors)

Definition of Factor functions

Definition:

Let $\boldsymbol{D}$ be a set of random variables. We define a factor f to be a function from $Val(D)$ to $\mathbb{R}$. A factor is non-negative if all its values are non-negative. The set of variables D is called the scope of the factor and denoted as $Scope(\boldsymbol{D})$.

$Val(D)$ represents the set of values D can take.

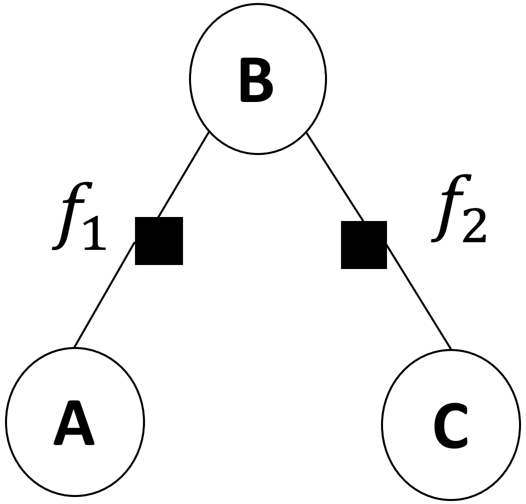
Example:

A	B	$f(A,B)$
a_0	b_0	30
a_0	b_1	5
a_1	b_0	1
a_1	b_1	10

$\boldsymbol{D} = \{A, B\}$
$A = \{a_0, a_1\}$
$B = \{b_0, b_1\}$

Product of Factor Functions in a Factor Graph

- In HMMs, we derived the joint distribution from the graph representation: $P(S_1, \ldots, S_n, E_1, \ldots, E_n) = P(S_1)P(E_1|S_1)\prod P(S_i|S_{i-1})P(E_i|S_i)$
- For a Factor Graph, the joint distribution can be derived from the product of factor functions (given that all factor functions are non-negative)



Example Factor Graph over variables A, B, C.

$$P(A, B, C) = \frac{1}{Z} f_1(A, B) f_2(B, C)$$

where, the normalization Z is given as

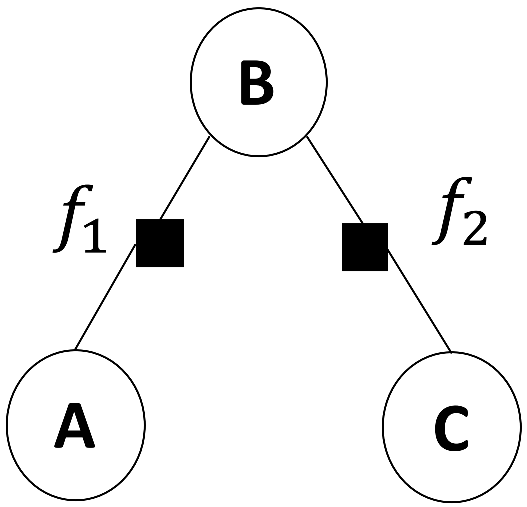
$$Z = \sum_{A,B,C} f(A, B, C) = \sum_{A,B,C} f_1(A, B) f_2(B, C)$$

Z is also referred to as the *partition function.*

Example of product of factor functions

Two factors f_1 and f_2 are multiplied in a way that "matches up" the common variables

$$f(A, B, C) = f_1(A, B) f_2(B, C)$$



f_1

a^1	b^1	0.5
a^1	b^2	0.8
a^2	b^1	0.1
a^2	b^2	0
a^3	b^1	0.3
a^3	b^2	0.9

f_2

b^1	c^1	0.5
b^1	c^2	0.7
b^2	c^1	0.1
b^2	c^2	0.2

f

a^1	b^1	c^1	$0.5 \cdot 0.5 = 0.25$
a^1	b^1	c^2	$0.5 \cdot 0.7 = 0.35$
a^1	b^2	c^1	$0.8 \cdot 0.1 = 0.08$
a^1	b^2	c^2	$0.8 \cdot 0.2 = 0.16$
a^2	b^1	c^1	$0.1 \cdot 0.5 = 0.05$
a^2	b^1	c^2	$0.1 \cdot 0.7 = 0.07$
a^2	b^2	c^1	$0 \cdot 0.1 = 0$
a^2	b^2	c^2	$0 \cdot 0.2 = 0$
a^3	b^1	c^1	$0.3 \cdot 0.5 = 0.15$
a^3	b^1	c^2	$0.3 \cdot 0.7 = 0.21$
a^3	b^2	c^1	$0.9 \cdot 0.1 = 0.09$
a^3	b^2	c^2	$0.9 \cdot 0.2 = 0.18$

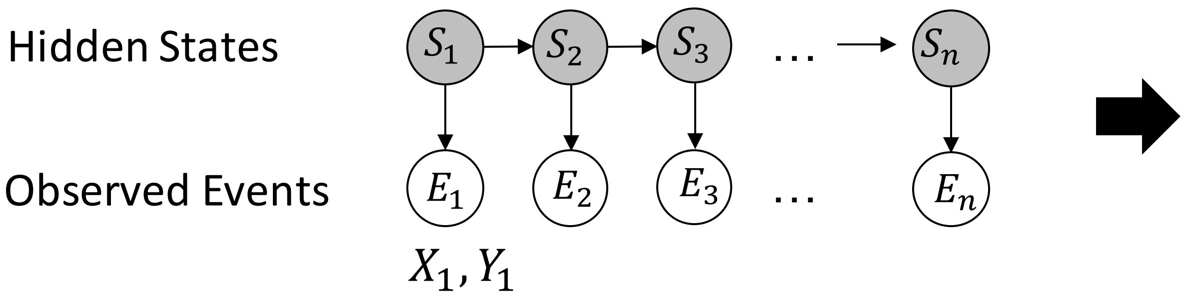
For example, $f(a^2, b^1, c^1) = f_1(a^2, b^1) f_2(b^1, c^1)$

Conversion of a Hidden Markov Model to a Factor Graph– Two dimension

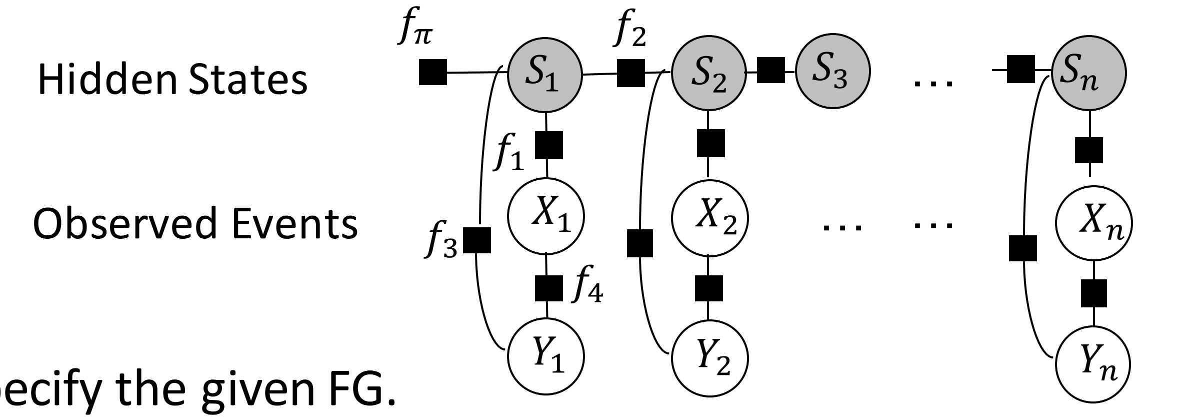
Assume that at each time point, two observations are made corresponding to random variables X and Y.

Example: Let $|\boldsymbol{S}| = \mathbf{10}, |\boldsymbol{X}| = \mathbf{10}, |\boldsymbol{Y}| = \mathbf{10}$

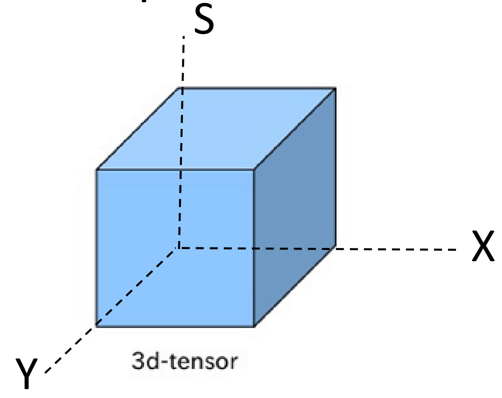
Hidden Markov Model



Factor Graph of the HMM

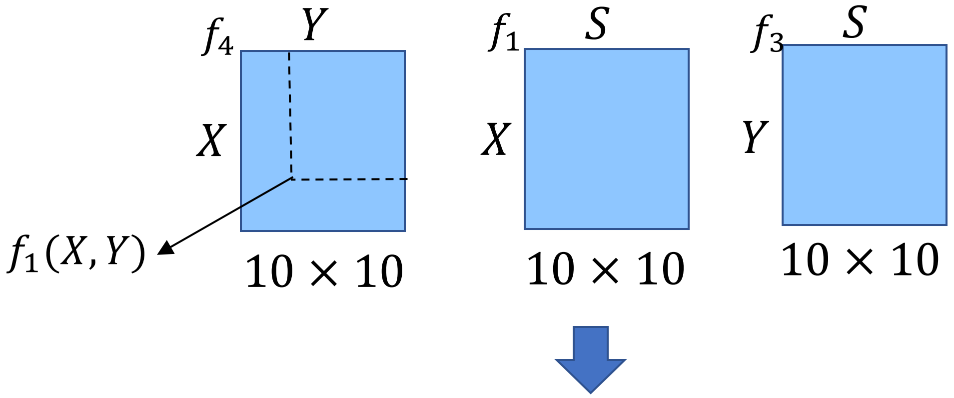


Fewer number of parameters are required are required to specify the given FG.



size of tensor is exponential

$10 \times 10 \times 10 = \mathbf{1000}$



size of three matrices (smaller exponent)

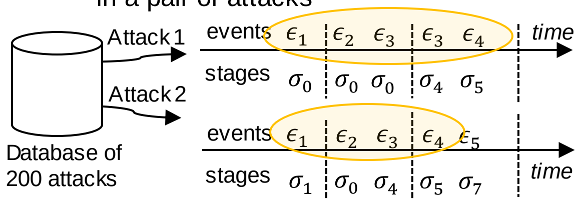
$10 \times 10 + 10 \times 10 + 10 \times 10 = \mathbf{300}$

Modeling the credential stealing attack using Factor Graphs - Data

State space of variables
Attack stage: $X = \{\sigma_0, \sigma_1, \dots, \sigma_7\}$
(Observed) Events: $E = \{\epsilon_1, \dots, \epsilon_5\}$

OFFLINE ANNOTATION ON PAST ATTACKS

a) Annotated events and attack stages in a pair of attacks



b) Event-stage annotation table for the attack pair (Attack 1 and Attack 2)

Event	Attack stage
$\{\epsilon_1\}$	$\{\sigma_0 \vert \sigma_1\}$
$\{\epsilon_2\}$	$\{\sigma_0\}$
$\{\epsilon_3\}$	$\{\sigma_4\}$
$\{\epsilon_4\}$	$\{\sigma_5\}$
$\{\epsilon_5\}$	$\{\sigma_7\}$

- Multi-stage credential stealing attack where the attack stage is not observed; however events which are related to the attack stage are observed
- Goal is to detect and pre-empt the attack
- **Model assumptions**
  - There are multivariate relationships among the events
  - There is no restriction on order of the relationships (can be non-causal or correlation based)

- Markov Model and Bayesian Networks cannot be used in this scenarios

- Factor graphs can be used for modeling highly complex attacks, where the causal relations among the events are not immediately clear.

ϵ_1	vulnerability scan	σ_0	benign
ϵ_2	login	σ_1	discovery
ϵ_3	sensitive_uri	σ_4	privilege escalation
ϵ_4	new_library	σ_5	persistence

Modeling the credential stealing attack using Factor Graphs

OFFLINE LEARNING OF FACTOR FUNCTIONS

Example patterns, stages, probabilities, and significance learned from the attack pair

Pattern	Attack stages	Probability in past attacks	Significance (p-value)
$[\epsilon_1, \epsilon_3, \epsilon_4]$	$[\sigma_1, \sigma_4, \sigma_5]$	q_a	p_a
$[\epsilon_1]$	$[\sigma_0 \mid \sigma_1]$	q_b	p_b

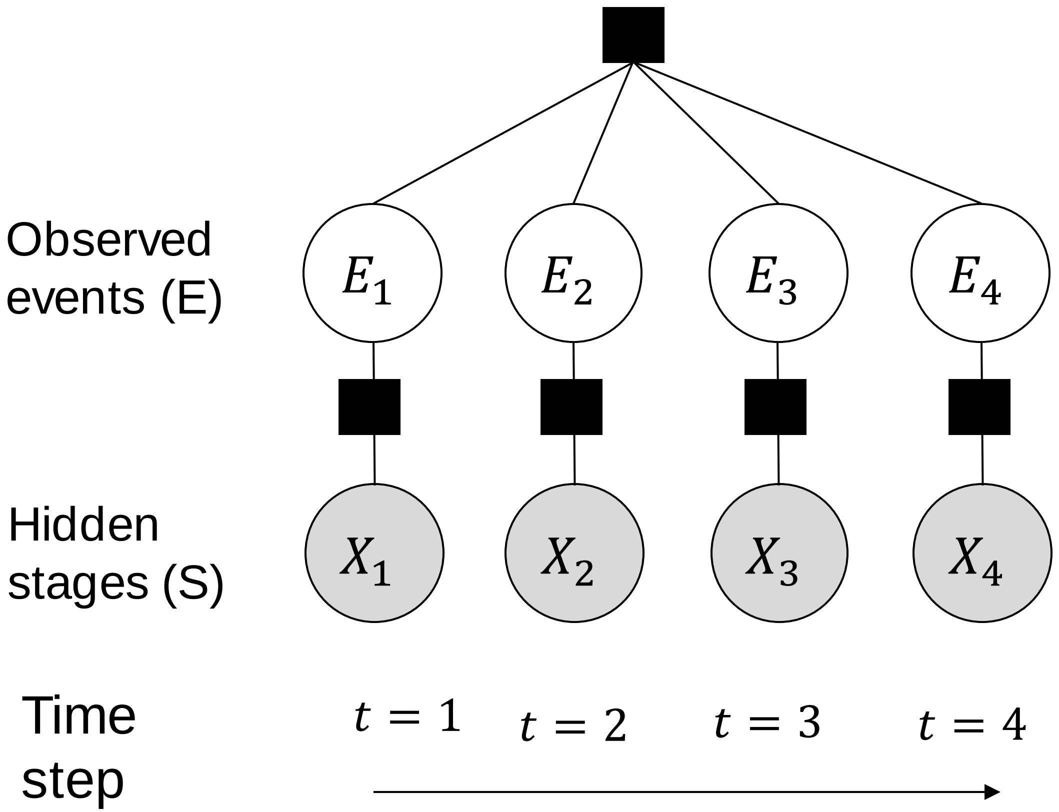
...

$$f(E) = \exp\{q_E(1 - p_E)\}$$

A factor function defined on the learned pattern, stages, and its significance

DETECTION OF UNSEEN ATTACKS

Factor Graph



Advantages and Disadvantages of Factor Graph

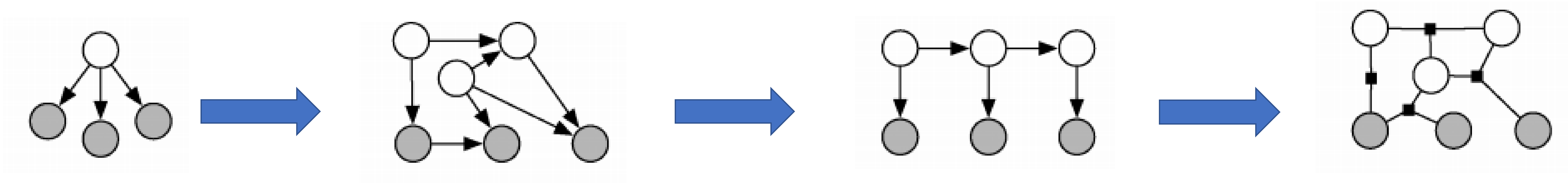
Advantage

- Factor graph subsumes HMMs, Markov Random Fields, Bayesian Networks etc.

Disadvantage

- If the problem is well represented by specific models such as Bayesian Networks, HMMs, Naïve Bayes or other graphical models then there is no need to go to generalize your problem as a factor graphs

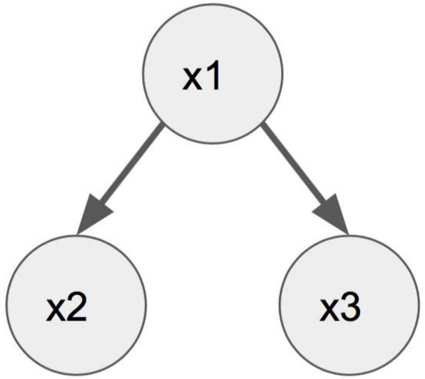
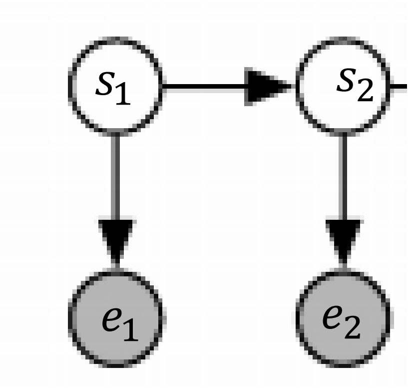
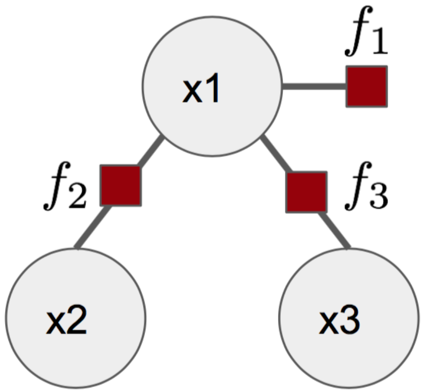
Taxonomy of graphical models



Naïve Bayes | Bayesian Network | Hidden Markov Model | Factor Graph

Conditional probabilities and statistical dependencies can be represented by a general type of graph: Factor Graph

Bayesian Networks vs. Hidden Markov Models vs. Factor Graphs

Bayesian Network	Hidden Markov Model	Factor Graph
		
$p(x_1)p(x_2\|x_1)p(x_3\|x_1)$	$p(s_1)p(e_1\|s_1)p(s_2\|s_1)p(e_2\|s_2)$	$\frac{1}{Z}f_1(x_1)f_2(x_2,x_1)f_3(x_1,x_3)$
Product of conditional probabilities	Product of Temporal dependencies among variable	Product of dependencies using univariate, bivariate, or multivariate functions
Causal relationships	Temporal and statistical dependencies	Both types of relations (including prior on a variable)