CS5340:
Tutorial 5

Asst. Prof. Harold Soh

TAs: Chen Kaiqi

Course Schedule

Week	Date	Lecture Topic	Tutorial Topic
1	12 Jan	Introduction to Uncertainty Modeling + Probability Basics	Introduction
2	19 Jan	Simple Probabilistic Models	Probability Basics
3	26 Jan	Bayesian networks (Directed graphical models)	More Basic Probability
4	2 Feb	Markov random Fields (Undirected graphical models)	DGM modelling and d-separation
5	9 Feb	Variable elimination and belief propagation	MRF + Sum/Max Product
6	16 Feb	Factor graph and the junction tree algorithm	**Quiz 1**
-	-	RECESS WEEK	
7	2 Mar	Mixture Models and Expectation Maximization (EM)	Linear Gaussian Models
8	9 Mar	Hidden Markov Models (HMM)	Probabilistic PCA
9	16 Mar	Monte-Carlo Inference (Sampling)	Linear Gaussian Dynamical System
10	23 Mar	Variational Inference	MCMC + Sequential VAE
11	30 Mar	Inference and Decision-Making (Special Topic)	**Quiz 2**
12	6 Apr	Gaussian Processes (Special Topic)	Wellness Day
13	13 Apr	**Project Presentations**	Closing

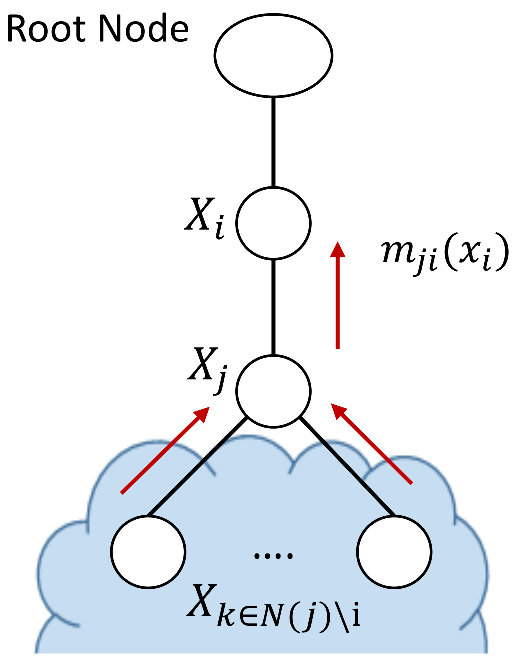
Sum-Product Algorithm

- Two phases:

$$m_{ji}(x_i) = \sum_{x_j} \left(\psi^E(x_j)\psi(x_i, x_j) \prod_{k \in \mathcal{N}(j)\backslash i} m_{kj}(x_j) \right)$$

1. Messages flow inward from leaves toward the root.

2. Initiated once all incoming messages have been received by the root node – messages flow outward from root toward the leaves.



Message-Passing Protocol

A node can send a message to a neighboring node when (and only when) it has received messages from all of its other neighbors.

Sum-Product Algorithm

$\text{SUM-PRODUCT}(\mathcal{T}, E)$ // main steps of the "Sum-Product Algorithm"
    $\text{EVIDENCE}(E)$
    $f = \text{CHOOSEROOT}(\mathcal{V})$
    for $e \in \mathcal{N}(f)$
        $\text{COLLECT}(f, e)$
    for $e \in \mathcal{N}(f)$
        $\text{DISTRIBUTE}(f, e)$
    for $i \in \mathcal{V}$
        $\text{COMPUTEMARGINAL}(i)$

$\text{EVIDENCE}(E)$ // add **evidence potentials** (convert conditioning into marginalization)
    for $i \in E$
        $\psi^E(x_i) = \psi(x_i)\delta(x_i, \bar{x}_i)$
    for $i \notin E$
        $\psi^E(x_i) = \psi(x_i)$

$\text{COLLECT}(i, j)$ // messages flow **inward** from leaves toward the root
    for $k \in \mathcal{N}(j) \backslash i$
        $\text{COLLECT}(j, k)$
    $\text{SENDMESSAGE}(j, i)$

$\text{DISTRIBUTE}(i, j)$ // messages flow **outward** from root toward the leaves
    $\text{SENDMESSAGE}(i, j)$
    for $k \in \mathcal{N}(j) \backslash i$
        $\text{DISTRIBUTE}(j, k)$

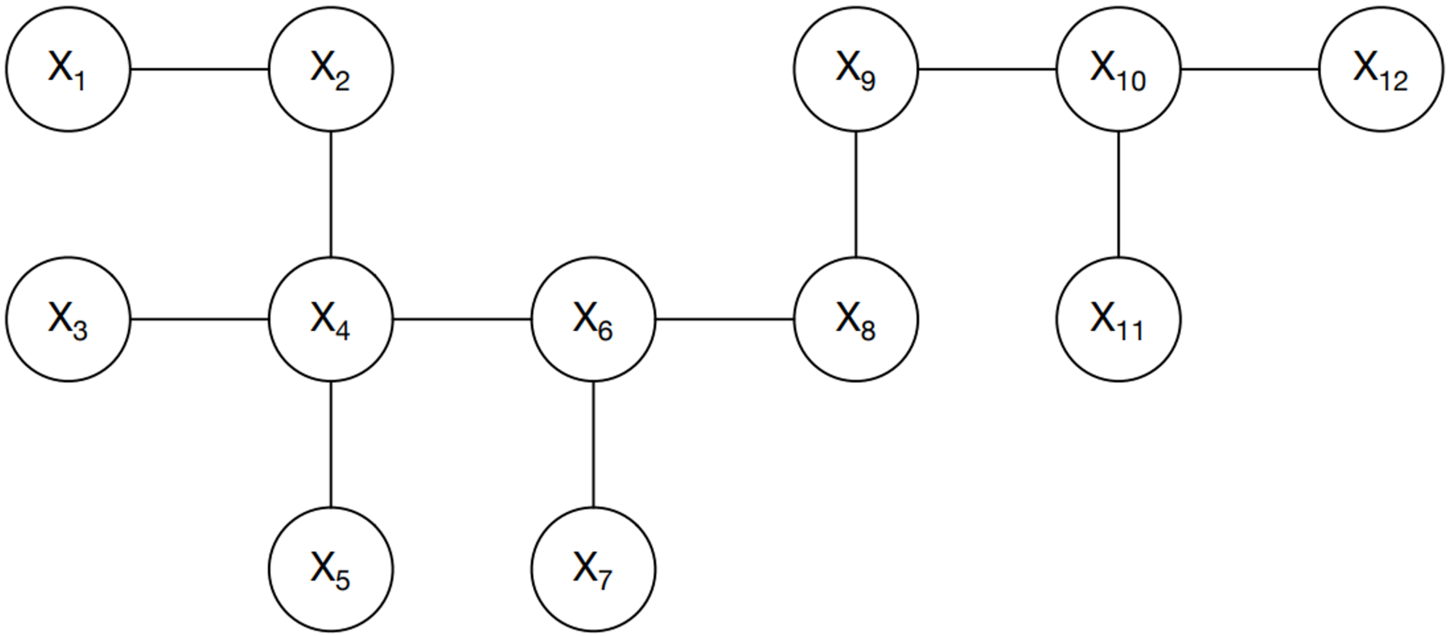
$\text{SENDMESSAGE}(j, i)$ // intermediate factors (messages)

$$m_{ji}(x_i) = \sum_{x_j} \left(\psi^E(x_j)\psi(x_i, x_j) \prod_{k \in \mathcal{N}(j) \backslash i} m_{kj}(x_j) \right)$$

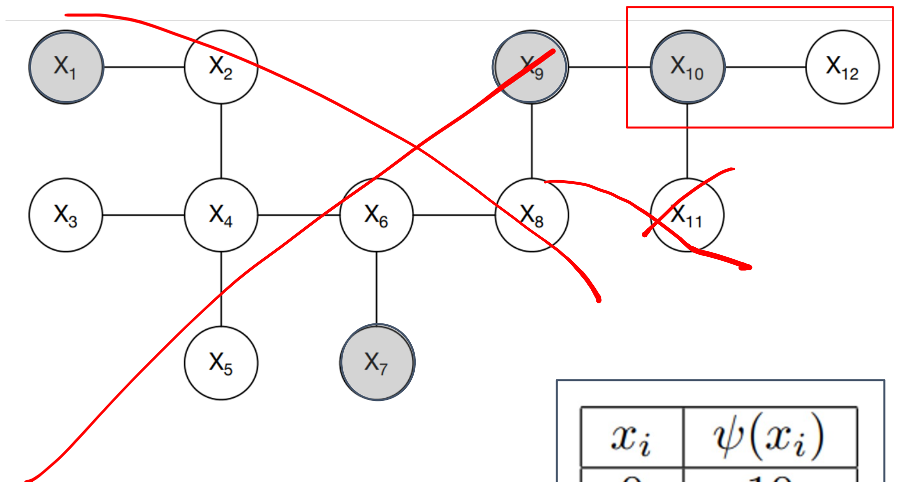
$\text{COMPUTEMARGINAL}(i)$ // message to final node

$$p(x_i) \propto \psi^E(x_i) \prod_{j \in \mathcal{N}(i)} m_{ji}(x_i)$$

MRT Inference (Again!)



Problem 1.a. Compute $p(x_{12} = 1 | x_1 = 0, x_7 = 0, x_9 = 1, x_{10} = 0)$.



$$p(x_{12}|x_1, x_7, x_9, x_{10}) = p(x_{12}|x_{10}) = \frac{p(x_{10}, x_{12})}{\sum_{x_{12}} p(x_{10}, x_{12})}$$

$$= \frac{m(x_{10})\psi(x_{10})\psi(x_{10}, x_{12})\psi(x_{12})}{\sum_{x_{12}} m(x_{10})\psi(x_{10})\psi(x_{10}, x_{12})\psi(x_{12})}$$

$$= \frac{\psi(x_{10}, x_{12})\psi(x_{12})}{\sum_{x_{12}} \psi(x_{10}, x_{12})\psi(x_{12})}$$

$$= \frac{\psi(x_{10} = 0, x_{12} = 1)\psi(x_{12} = 1)}{\sum_{x_{12}} \psi(x_{10} = 0, x_{12})\psi(x_{12})}$$

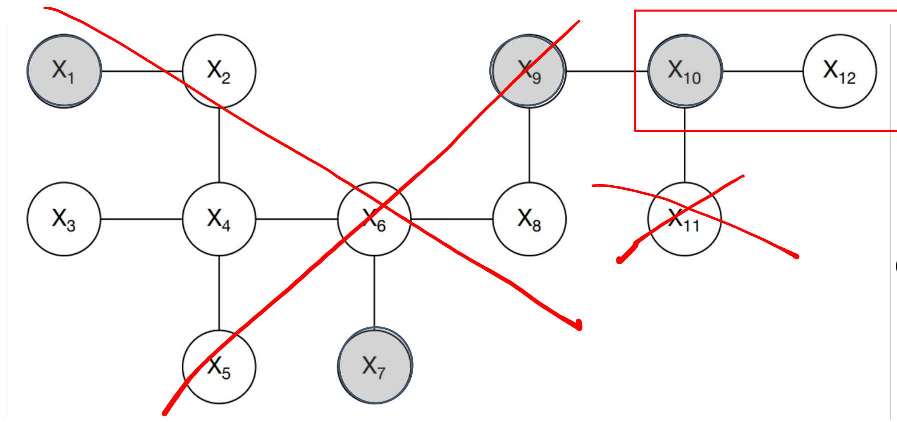
$$= \frac{5 \times 2}{20 \times 10 + 5 \times 2} = \frac{1}{12} = 0.0476$$

x_i	$\psi(x_i)$
0	10
1	2

x_i	x_j	$\psi(x_i, x_j)$
0	0	20
0	1	5
1	0	5
1	1	20

Sum-Product approach:

Problem 1.a. Compute $p(x_{12} = 1 | x_1 = 0, x_7 = 0, x_9 = 1, x_{10} = 0)$.



$$m_{i\to j}(x_j) = \sum_{x_i \in \{0,1\}} \left(\psi^E(x_i)\psi(x_i, x_j) \cancel{\prod_{x_k \in \text{neighbors}(x_i)\setminus x_j} m_{k\to i}(x_i)} \right)$$

$\psi^E(x_i) = \delta(x_i = \hat{x}_i)\psi(x_i)$ if $x_i \in E$ and $\psi^E(x_i) = \psi(x_i)$ otherwise

$$m_{ji}(x_i) = \sum_{x_j} \left(\psi^E(x_j)\psi(x_i, x_j) \prod_{k \in \mathcal{N}(j)\setminus i} m_{kj}(x_j) \right)$$

$$p(x_f \,|\, \bar{x}_E) \propto \psi^E(x_f) \prod_{e \in \mathcal{N}(f)} m_{ef}(x_f)$$

x_i	$\psi(x_i)$
0	10
1	2

x_i	x_j	$\psi(x_i, x_j)$
0	0	20
0	1	5
1	0	5
1	1	20

Compute the message from x_{10} to x_{12} :

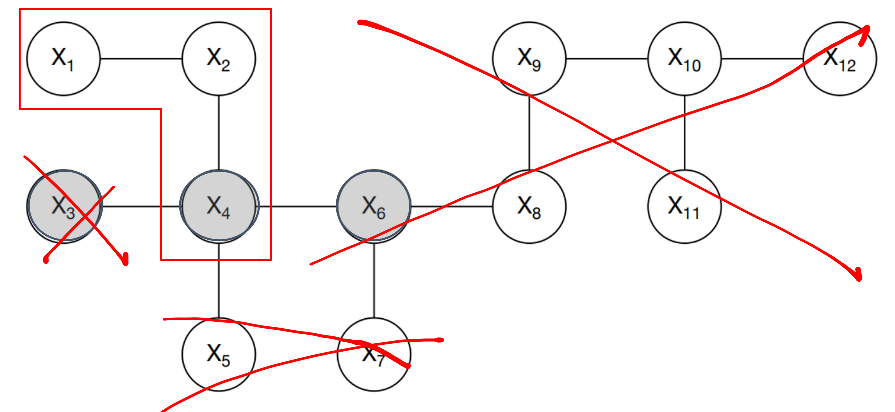
$m_{x_{10}\to x_{12}}$		
$x_{12} = 0$	$10 \times 20 + 0 \times 5$	200
$x_{12} = 1$	$10 \times 5 + 0 \times 20$	50

$$\tilde{p}(x_{12} = \hat{x}_{12} | x_1 = 0, x_7 = 0, x_9 = 1, x_{10} = 0) = \psi(x_{12} = \hat{x}_{12}) \times m_{x_{10}\to x_{12}}(x_{12} = \hat{x}_{12})$$

$\tilde{p}$		
$x_{12} = 0$	10×200	2000
$x_{12} = 1$	2×50	100

$$p(x_{12} = 1 | x_1 = 0, x_7 = 0, x_9 = 1, x_{10} = 0) = \frac{100}{2000 + 100} = \frac{1}{21} = 0.0476$$

Problem 1.b. Compute $p(x_1 = 1 | x_3 = 0, x_4 = 1, x_6 = 0)$.



$$p(x_1|x_3,x_4,x_6) = p(x_1|x_4) = \frac{p(x_1,x_4)}{\sum_{x_1} p(x_1,x_4)}$$

$$= \frac{\sum_{x_2} \psi(x_1,x_2)\psi(x_2)\psi(x_2,x_4)\psi(x_1)\psi(x_4)m(x_4)}{\sum_{x_1}\sum_{x_2} \psi(x_1,x_2)\psi(x_2)\psi(x_2,x_4)\psi(x_1)\psi(x_4)m(x_4)}$$

$$= \frac{\sum_{x_2} \psi(x_1,x_2)\psi(x_2)\psi(x_2,x_4)\psi(x_1)}{\sum_{x_1}\sum_{x_2} \psi(x_1,x_2)\psi(x_2)\psi(x_2,x_4)\psi(x_1)}$$

x_i	$\psi(x_i)$
0	10
1	2

$$= \frac{\sum_{x_2} \psi(x_1=1,x_2)\psi(x_2)\psi(x_2,x_4=1)\psi(x_1=1)}{\sum_{x_1}\sum_{x_2} \psi(x_1,x_2)\psi(x_2)\psi(x_2,x_4=1)\psi(x_1)}$$

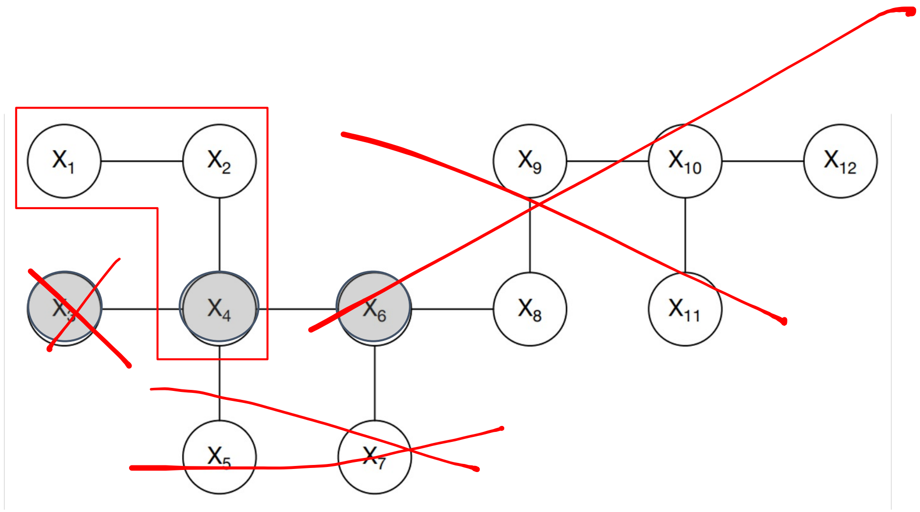
$$= \frac{5\times 10\times 5\times 2 + 20\times 2\times 20\times 2}{20\times 10\times 5\times 10 + 5\times 2\times 20\times 10 + 5\times 10\times 5\times 2 + 20\times 2\times 20\times 2}$$

x_i	x_j	$\psi(x_i,x_j)$
0	0	20
0	1	5
1	0	5
1	1	20

$$= \frac{7}{47} = 0.1489$$

Sum-Product approach:

Problem 1.b. Compute $p(x_1 = 1 | x_3 = 0, x_4 = 1, x_6 = 0)$.



Compute the message from x_4 to x_2 :

$m_{x_4 \to x_2}$		
$x_2 = 0$	$0 \times 20 + 2 \times 5$	10
$x_2 = 1$	$0 \times 5 + 2 \times 20$	40

Next, compute the message from x_2 to x_1 :

$m_{x_2 \to x_1}$		
$x_1 = 0$	$10 \times 20 \times 10 + 2 \times 5 \times 40$	2400
$x_1 = 1$	$10 \times 5 \times 10 + 2 \times 20 \times 40$	2100

$$m_{ji}(x_i) = \sum_{x_j} \left(\psi^E(x_j) \psi(x_i, x_j) \prod_{k \in \mathcal{N}(j) \setminus i} m_{kj}(x_j) \right)$$

$$p(x_f \,|\, \bar{x}_E) \propto \psi^E(x_f) \prod_{e \in \mathcal{N}(f)} m_{ef}(x_f)$$

$$\tilde{p}(x_1 = \hat{x}_1 | x_3 = 0, x_4 = 1, x_6 = 0) = \psi(x_1 = \hat{x}_1) \times m_{x_2 \to x_1}(x_1 = \hat{x}_1)$$

$\tilde{p}$		
$x_1 = 0$	10×2400	24000
$x_1 = 1$	2×2100	4200

x_i	$\psi(x_i)$
0	10
1	2

x_i	x_j	$\psi(x_i, x_j)$
0	0	20
0	1	5
1	0	5
1	1	20

$$p(x_1 = 1 | x_3 = 0, x_4 = 1, x_6 = 0) = \frac{4200}{24000 + 4200} = \frac{7}{47} = 0.1489$$

Linear Gaussian Model (Gaussian Bayesian Network)

Problem 3.a. We will build our way up towards this model. As a prelude, consider K *independent* univariate Gaussian random variables $x_1, x_2, \ldots, x_K$,

$$p(x_k) = \mathcal{N}(\mu_k, \sigma_k^2)$$

for $k = 1, 2, \ldots, K$. Define the random variable x_L,

$$x_L = b + \sigma_L \epsilon + \sum_{k=1}^{K} w_k x_k$$

where $\epsilon \sim \mathcal{N}(0, 1)$.

1. Draw out the DGM for the model described above.
2. Show that $p(x_L | x_1, \ldots, x_K) = \mathcal{N}\left(b + \sum_{k=1}^{K} w_k x_k, \sigma_L^2\right)$. In other words, x_L is Gaussian distributed with mean $b + \sum_{k=1}^{K} w_k x_k$ and variance σ_L^2.
3. Define the random variable $\mathbf{x} = (x_1, x_2, \ldots, x_K, x_L)$. Show that $\mathbf{x}$ is a *multivariate* Gaussian random variable. *Hint: Consider the definition of the multivariate Gaussian and the properties of Gaussians.*

Univariate Normal Distribution

- Also known as the Gaussian distribution.
- Univariate normal distribution describes single continuous variable X, i.e. $x \in \mathbb{R}$.
- Two parameters $\mu \in \mathbb{R}$ (mean) and $\sigma^2 > 0$ (variance).

Carl Friedrich Gauss
1777–1855

$$p(X = a \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \frac{(a-\mu)^2}{2\sigma^2}, \quad a \in \mathbb{R}$$

Or

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \frac{(x-\mu)^2}{2\sigma^2}$$

$$p(x) = \text{Norm}_x[\mu, \sigma^2]$$

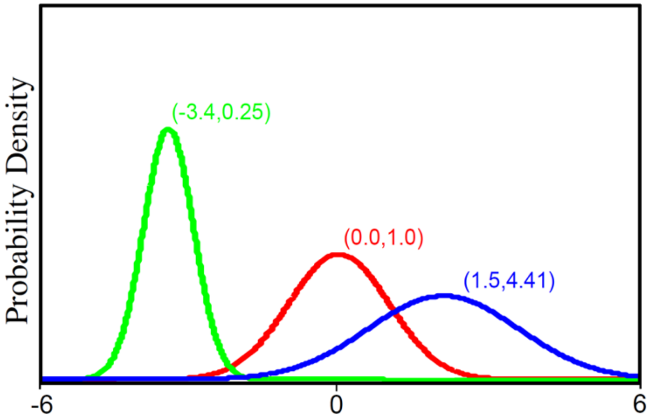


Image sources: “Pattern Recognition and Machine Learning”, Christopher Bishop
“Computer Vision: Models, Learning, and Inference”, Simon Prince

Multivariate Normal Distribution

- Multivariate normal distribution describes a *D*-dimensional continuous variable $\boldsymbol{X}$, i.e. $\boldsymbol{x} \in \mathbb{R}^D$.

- *D*-dimensional mean $\boldsymbol{\mu} \in \mathbb{R}^D$, and $D \times D$ symmetric positive definite covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}_+^{D \times D}$.

$$p(\boldsymbol{X} = \boldsymbol{a} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\{ -0.5(\boldsymbol{a} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{a} - \boldsymbol{\mu}) \}, \quad \boldsymbol{a} \in \mathbb{R}^D$$

Or

$$\boxed{\begin{aligned} p(\boldsymbol{x}) &= \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\{ -0.5(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \} \\ p(\boldsymbol{x}) &= \mathrm{Norm}_{\boldsymbol{x}}[\boldsymbol{\mu}, \boldsymbol{\Sigma}] \end{aligned}}$$