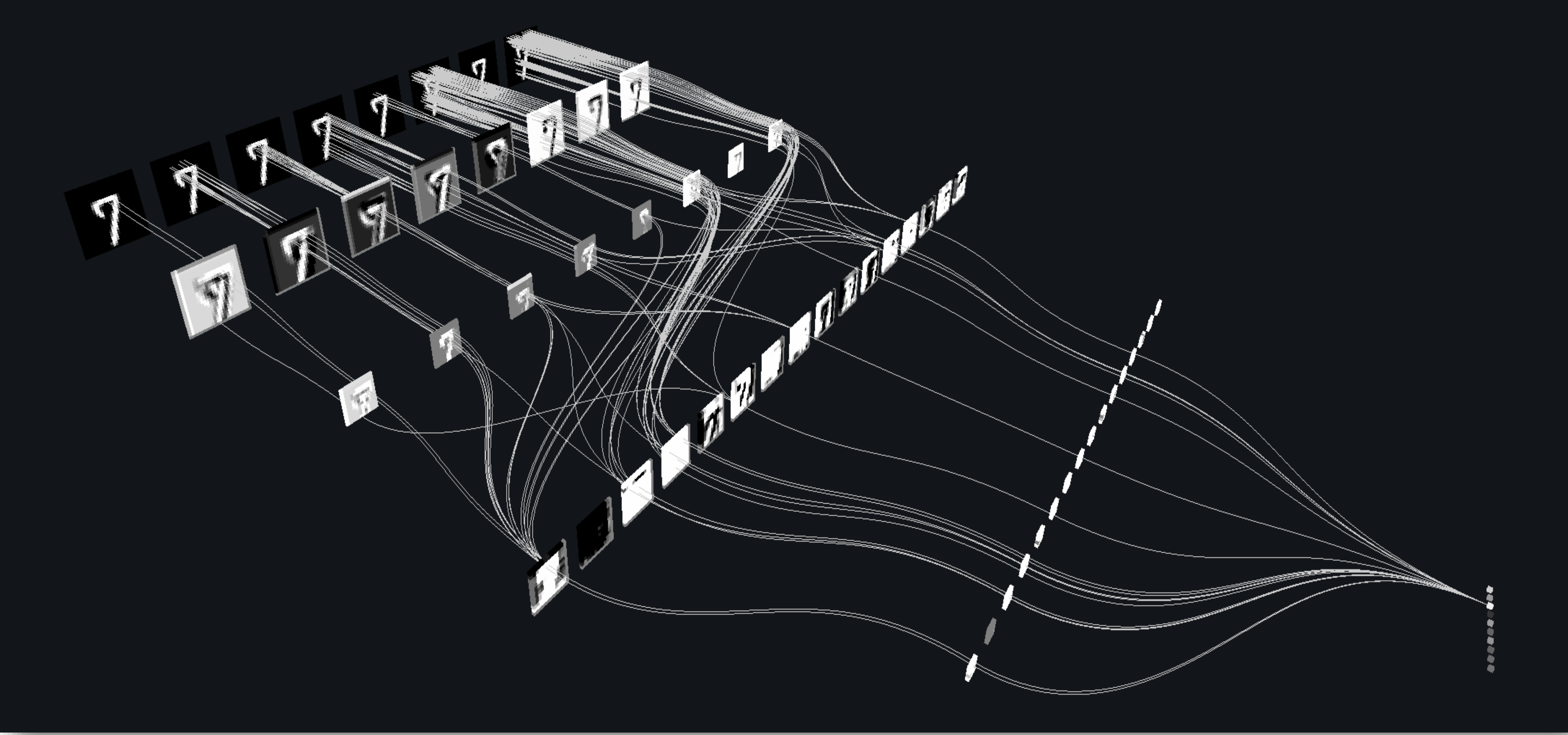
Deep Learning

Md. Jalil Piran, PhD
Asst. Professor
Computer Science and Engineering
Sejong University
Spring, 2021

Outline

- Introduction to Deep Learning (DL)
- The History of DL
- Programming Tools
- Artificial Neural Networks (ANNs)
- Optimization in DL
- Convolutional Neural networks (CNNs)
- **Unsupervised Pre-trained Networks (UPNs)**



ARCHITECTURE OF DEEP LEARNING

DL Architectures

- **Higher-level Architecture**
  - **Convolutional Neural Networks (CNNs)**
  - **Unsupervised Pre-trained Networks (UPNs)**
    - Autoencoders
    - Deep belief networks (DBNs)
    - Generative adversarial networks (GANs)
  - **Recurrent Neural Networks (RNNs)**
    - Bidirectional recurrent neural networks (BRNN)
    - LSTM
  - **Recursive Neural Networks**

Unsupervised Learning

Motivation and Strengths:

- Unsupervised learning is **not expensive** and **time consuming** like supervised learning.
- Unsupervised learning requires **no human intervention**.
- Unlabeled data is **easy** to **find** with large quantities, unlike labeled data which is scarce.

- **Weaknesses**

- More difficult than supervised learning because there is NO **Single objective** (like test set accuracy)

Unsupervised Feature Learning

- Train representations with unlabeled data.
  - Minimize an *unsupervised* training loss.
    - Often based on generic priors about characteristics of good features
    - Usually train 1 layer of features at a time.

UPNs

- Unsupervised pre-trained networks (UPNs)
  - Motivation: representation **leaning** and **transfer learning**
  - Deep belief networks (DBNs)
  - Autoencoders
  - Generative adversarial networks (GANs)
  - RBM

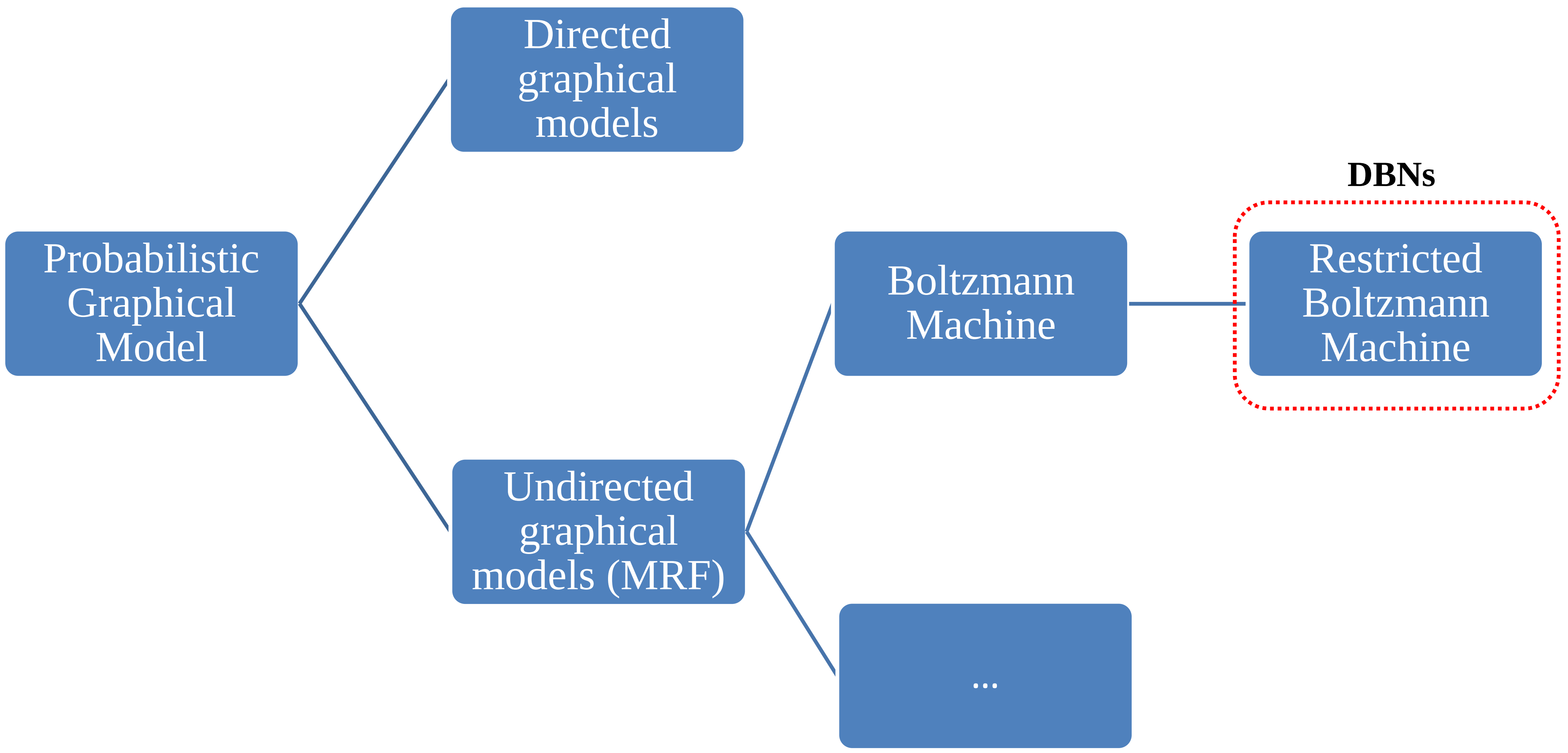
DBN

- DBN’s pre-requesitions
  - MRF
  - Sampling
  - RBMs

Restricted Boltzmann Machine (RBM)

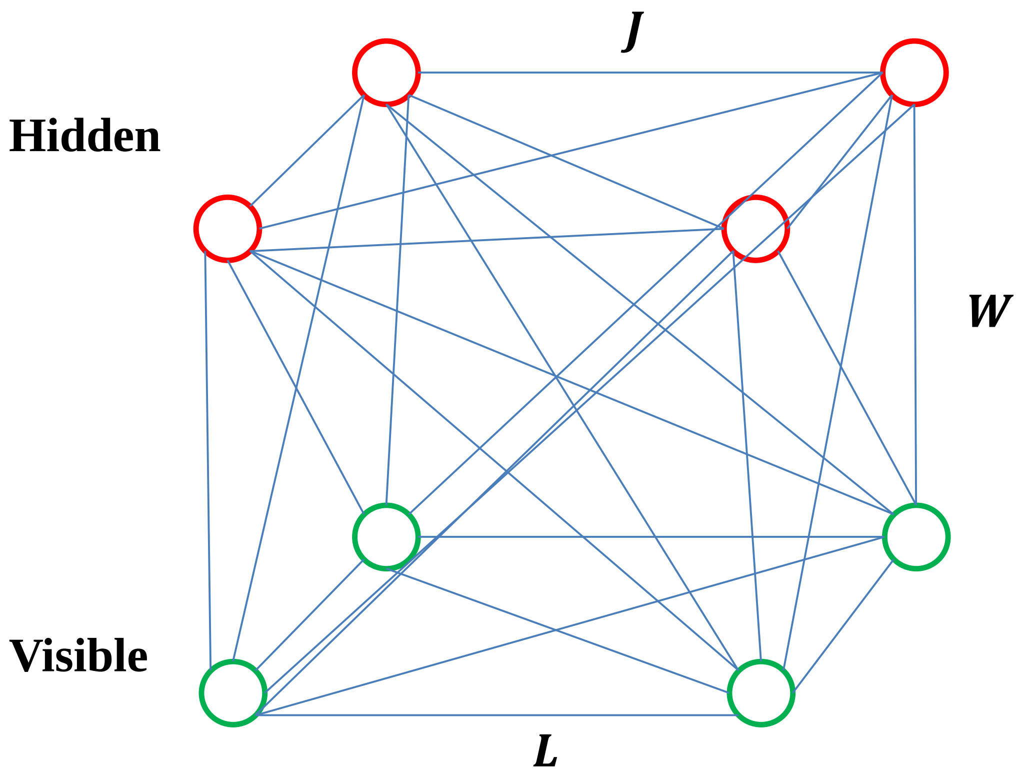
- RBMs are building blocks for the multi-layer leaning architectures, e.g. DBNs.
- RBMs are a special case of general **Boltzmann Machines** (BMs).
- BMs are a particular form of **Markov Random Field (MRF)**, a.k.a. **Markov networks** or **undirected graph models**.

DBNs


Directed graphical models
DBNs
Probabilistic Graphical Model
Boltzmann Machine
Restricted Boltzmann Machine
Undirected graphical models (MRF)
…

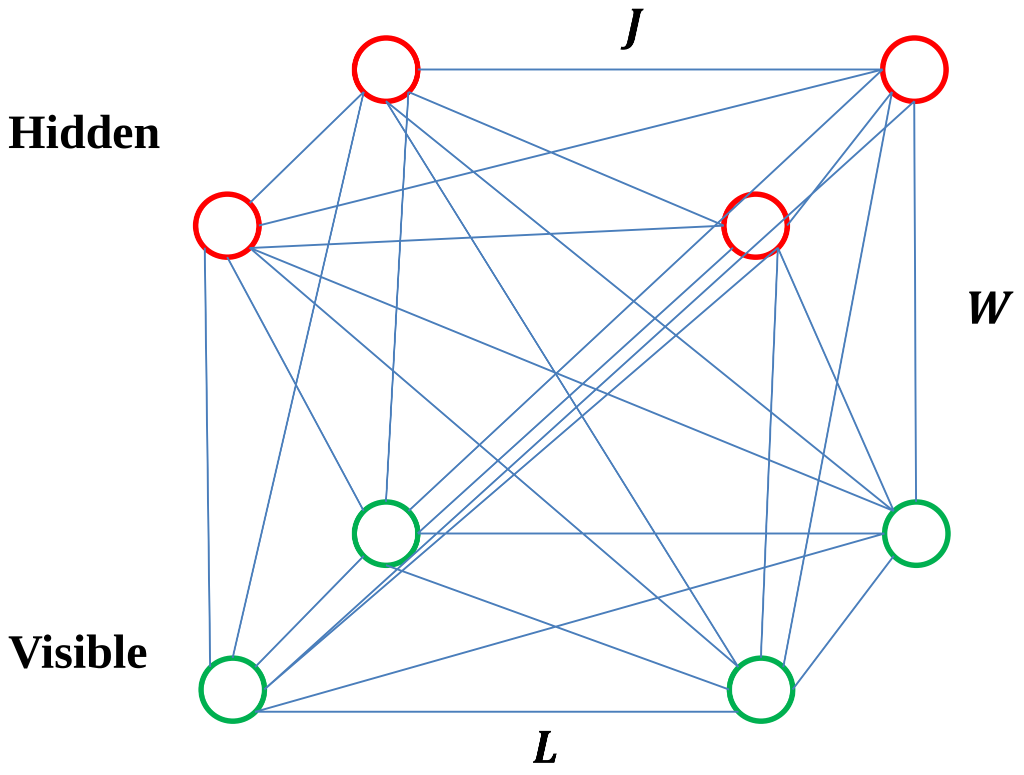
Boltzmann Machine

- BM: a network of symmetrically coupled **stochastic binary units**, 0s and 1s.
- Represents a **probability distribution**
- Can be used to **learn** important aspects of an unknown target distribution based on samples from this target distribution.
- Components:
  - **Visible units** $v \in \{0,1\}^D$
  - **Hidden units** $h \in \{0,1\}^D$
- Training process:
  - Adjusting the parameters to fit the training data.
  - Parameters: weights of the links between the nodes.

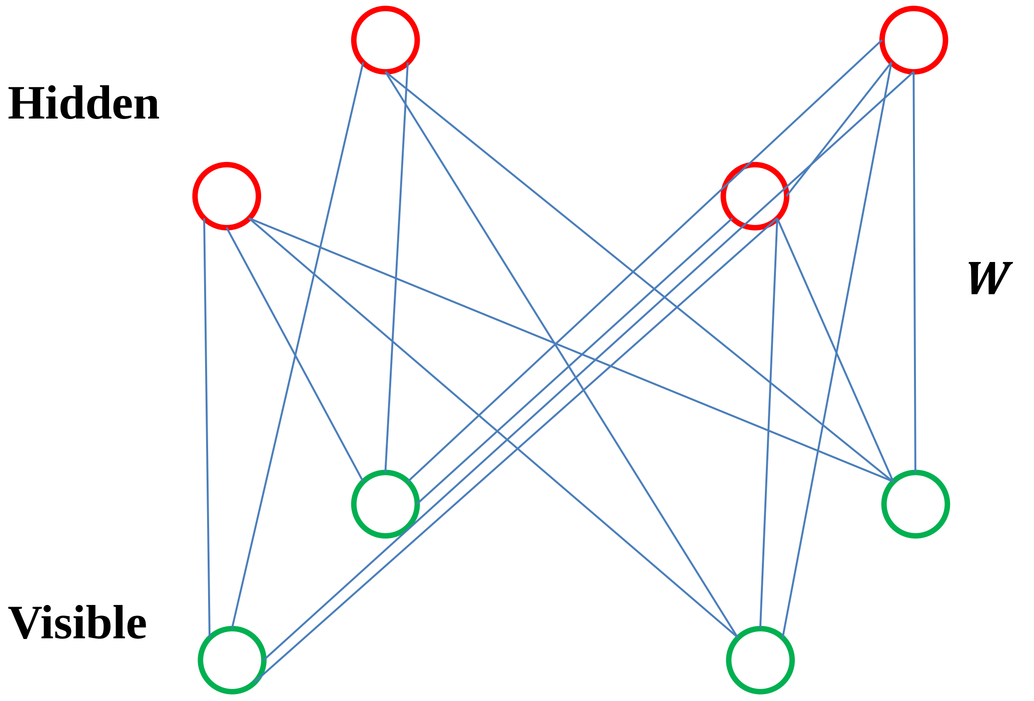


BM

- Training a BM is computationally difficult.
- Learning problem can be simplified by imposing **restriction** on the network topology.
  - e.g. remove the connections between the neurons in the same layer.



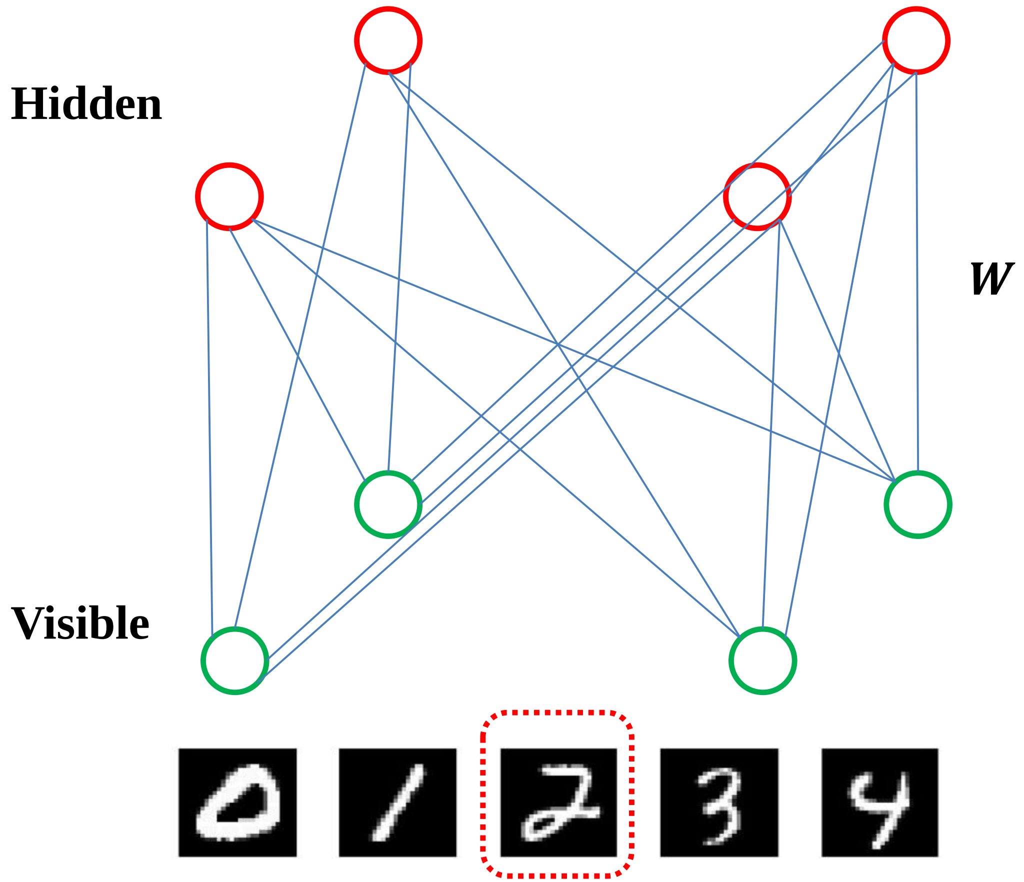
General Boltzmann Machine



Restricted Boltzmann Machine

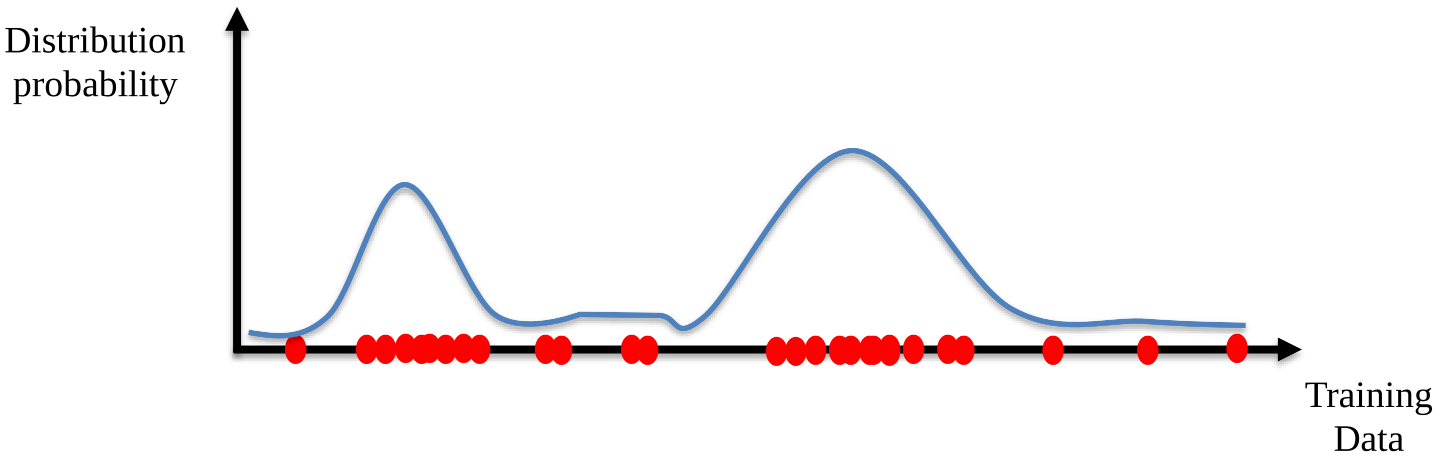
RBM

- Hidden units model the **dependencies** between the components of observations
  - e.g. non-linear feature detectors



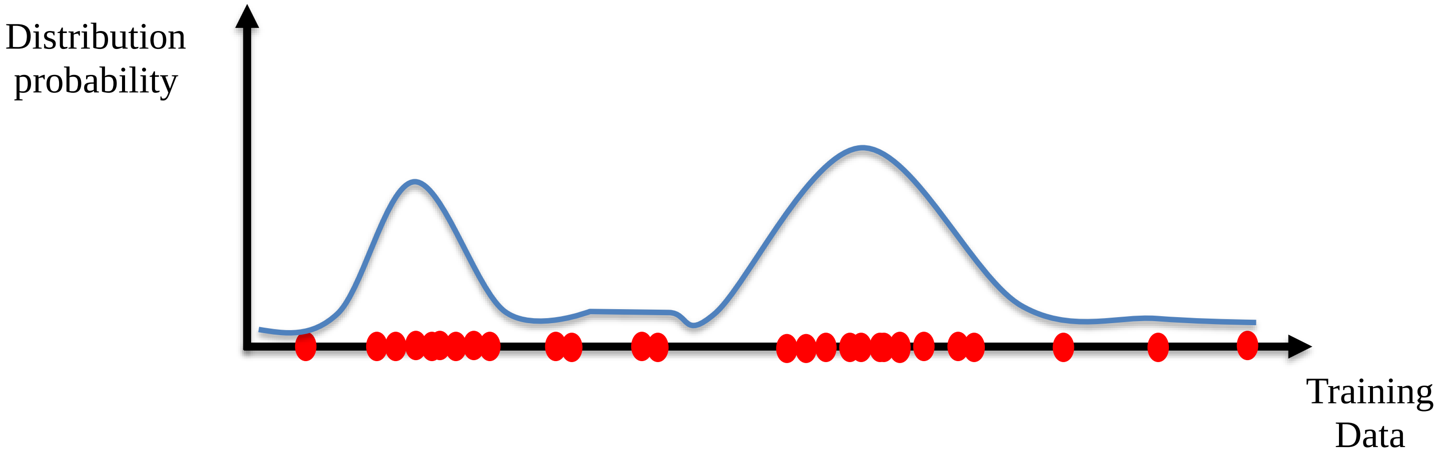
RBM

- Example
  - Just consider 1D data.



RBM

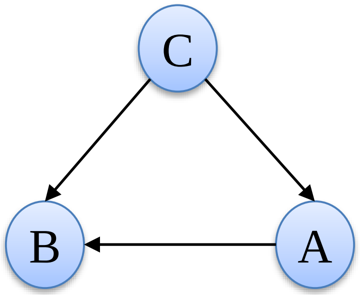
- Training
  - Usually based on gradient-based maximization of the likelihood.
  - Computing the gradient is computationally expensive
  - Sampling-based methods are employed to **approximate** the likelihood gradient
    - Sampling from an MRF is not straightforward
    - Solution: Markov Chain Monte Carlo (MCMC) methods easily applicable in the form of **Gibbs sampling.**



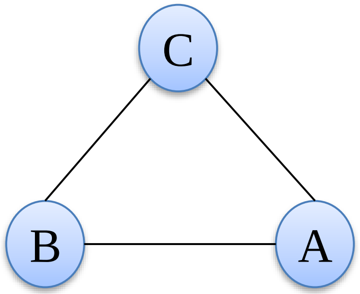
MRF

- Probabilistic Graphical Models

- **Directed graphical models**
  - *Bayesian networks*



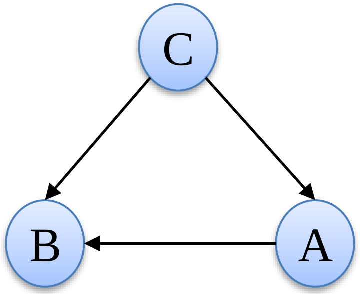
- **Undirected graphical models**
  - *MRF*



Graphs

- Directed graphical model

$$p(a,b,c) = p(c|a,b)P(a|b)p(C)$$
$$= p(c|a,b)p(b|a)p(a)$$



- **Definition**: for a graph with K nodes, the joint distribution is:

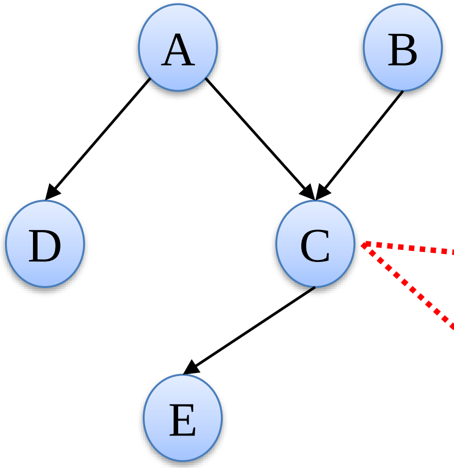
$$p(x_1, K, x_K) = \prod_{k=1}^{K} p(x_k|pa_k)$$

  - pa_k: the set of parents of x_k

MRF

- Both directed and undirected graphical models specify a factorization, e.g. how to express **the joint distribution**.
- Both define a set of conditional independence properties.

- **Directed graphical models**

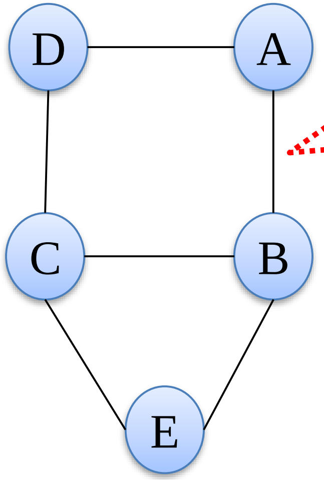


$P(C\|A,B)$			
A	B	0	1
0	0	0.9	0.1
0	1	0.2	0.8
1	0	0.9	0.1
1	1	0.01	0.99

$$P(A,B,C,D,E) = P(A)P(B)P(C|A,B)P(D|A)P(E|C)$$

- Parent-child
- Local conditional distribution

- **Undirected graphical models**

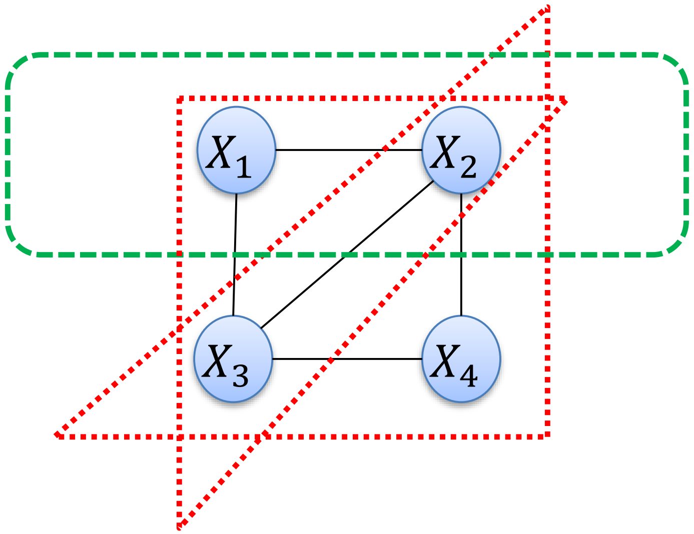


A	B	$\pi_1(A,B)$
0	0	1.0
0	1	0.5
1	0	0.5
1	1	2.0

$$P(A,B,C,D,E) = \frac{1}{Z}\pi_1(A,B)\pi_2(B,C,E)\pi_3(C,D)\pi_4(A,D)$$

- Maximal clique
- Potential function

- **Undirected graphical model**
  - **Clique**: a subset of the fully connected nodes.
  - **Maximal clique**: no node can be added such that the resulting set is still a clique



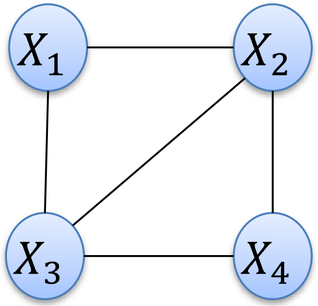
- We can define the factors in decomposition of the joint distribution as functions of the variable in the clique.

Graphs

- **Undirected factorization**
  - **Factorization**: $\psi_c(x_c)$ is a non-negative potential function of a maximal clique
  - **Partition function**: the normalization constant (Z)

$$p(\underline{x}) = \frac{1}{Z}\prod_c \psi_c(x_c)$$

- Example:



$$p(x_1, \dots, x_4) = \frac{1}{Z}\psi\overbrace{\boxed{(x_1, x_2, x_3)}}^{\text{Maximal clique 1}}\psi\overbrace{\boxed{(x_2, x_3, x_4)}}^{\text{Maximal clique 2}}$$

- Potential functions in exponential form:

$$\psi_c(x_c) = \exp\{-E(x_c)\}$$

Undirected Graphical Models

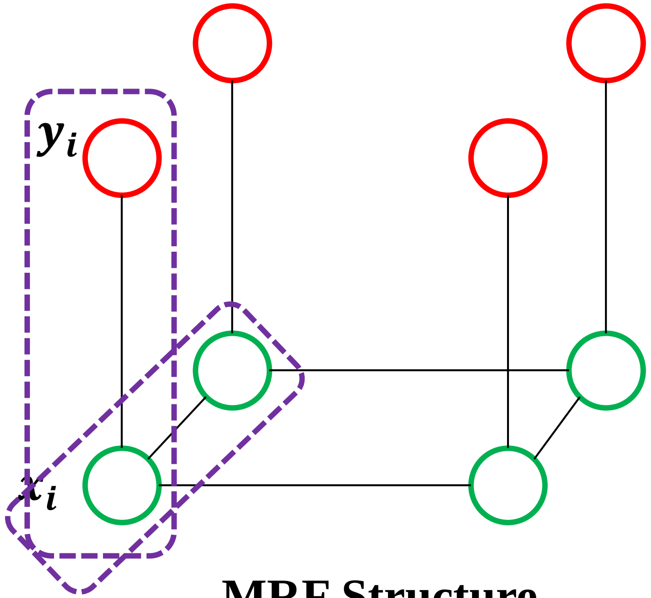
- Image de-noising example
- Goal: construct an MRF using these **prior knowledge**:
  - Flipping pixel color probability is 10%
  - $(y_i s)$: array of noisy pixels (noisy image)
  - $(x_i s)$: inferred original pixels (clean image)
  - y_i and x_i are strongly correlated, since noise level is small.
  - Neighboring pixels x_i and $x_i s$ in image are strongly correlated.
    - e.g. neighboring pixels have the same value
  - Bias toward $x_i = 0$

$\boldsymbol{y_i}, \{\boldsymbol{i} \in \mathbf{1..}\boldsymbol{D}\}$ $\boldsymbol{x_i}, \{\boldsymbol{i} \in \mathbf{1..}\boldsymbol{D}\}$

Undirected Graphical Models

- Image de-noising example
- Model as a pairwise MRF
- The graph has two types of cliques, each contains two variables (a pairwise MRF)
  - $\{x_i, y_i\}$ and $\{x_i, x_j\}$
  - $x_i \in \{-1,1\}$ and $y_i \in \{-1,1\}$
- $-\eta x_i y_i \qquad \eta > 0$
- $h x_i - \beta x_i x_j \qquad h, \beta > 0$
- $E(X,Y) = h\sum_i x_i - \beta \sum_{\{i,j\}} x_i x_j - \eta \sum_i x_i y_i$

$\boldsymbol{y_i, \{i \in 1..D\}}$ $\boldsymbol{x_i, \{i \in 1..D\}}$

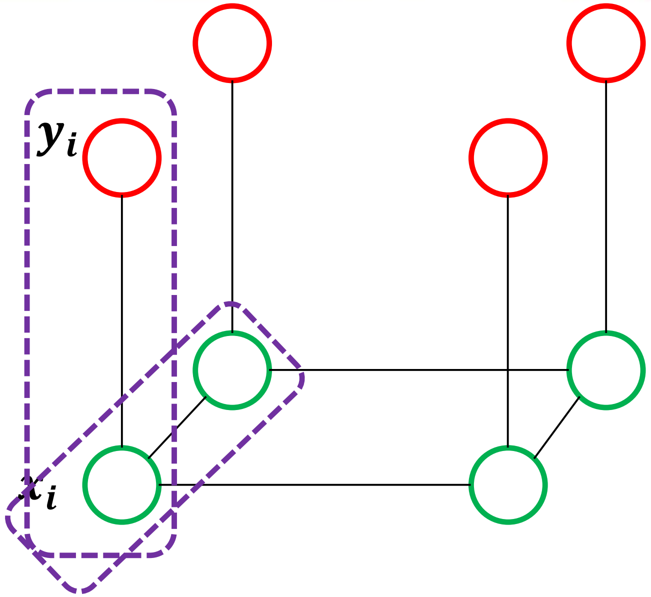


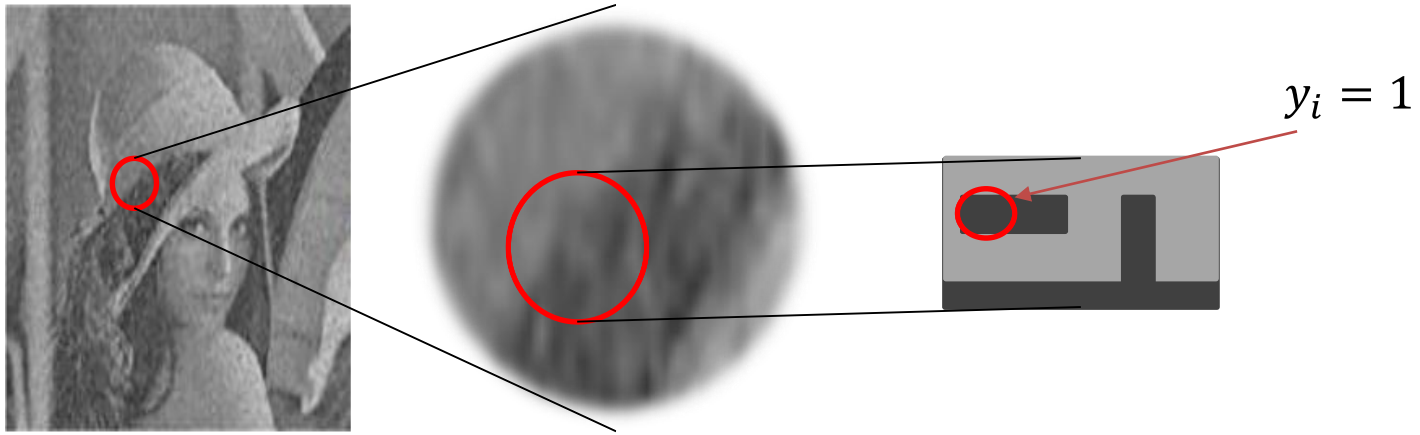
MRF Structure

Undirected Graphical Models

- **Inference using Iterated conditional modes (ICM)**
- $E(X,Y) = h\sum_i x_i - \beta\sum_{\{i,j\}} x_i x_j - \eta\sum_i x_i y_i$
  - Suppose: $h = 0, \beta = 1.0, \eta = 2.1$
- $E(x_i = -1) = [0 * -1] - 1.0[(-1 * -1) + (-1 * 1) + (-1 * 1) +$

Neighboring pixels



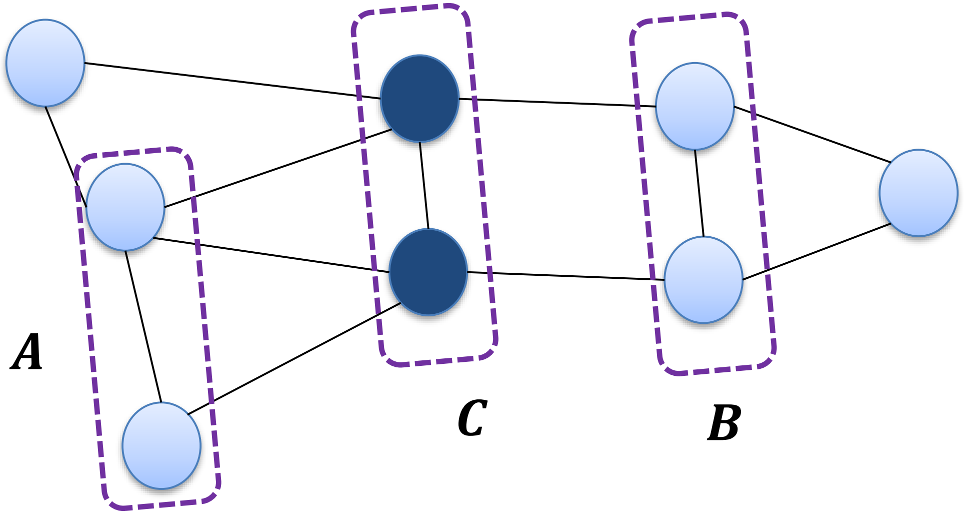


Undirected Graphical Models

- **Conditionally independent**
- $Nodes\ of\ set\ A\ and\ B\ are\ sepearted\ by\ the\ third\ set\ C$
- A and B are conditionally independent

$$\mathrm{A} \perp\!\!\!\perp \mathrm{B}|C$$

$$p(a_1, b_1|c_1, c_2) = p(a_1|c_1, c_2)p(b_1|c_1, c_2)$$



Office: #432, Dayang AI Center, **Phone**: 010-8999-8586, **email**: *piran@sejong.ac.kr*