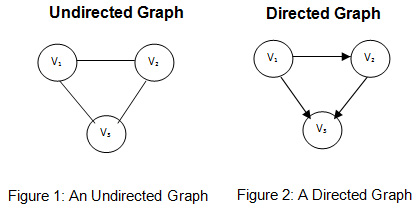
**Lecture 8 – Whiteboard notes**

## **Probabilistic Graphical models**

* Probability theory
  + Deals with random variable X with uncertainty
* Graph theory
  + Reduces complexity
  + Types of graphs:

[](https://www.google.com/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=0ahUKEwjd9KCI36LLAhXJu4MKHeUEDhoQjRwIBw&url=http://www.differencebetween.com/difference-between-directed-and-vs-undirected-graph/&psig=AFQjCNHeqQrbDq0qbp9CKGb2B5sJiCBZCA&ust=1457033526559449)

V1, V2, V3 are nodes. The lines connecting the nodes are edges/links

## **Types of Probabilistic Graphical models**

* Directed graph models:
  + Bayesian networks
* Undirected graphical models
  + Markov random fields

## **Probability theory**

* Sum Rule
* Product Rule

## **Bayesian network Concepts**

* Independence and inferences are two important concepts:
* Independence:
  + A and B are conditionally independent given C.
* Inference:
  + Estimate likelihood of one event based on other events

## **Bayesian networks**

* Follows baye’s rule
  + Example:
  + Also called Belief networks
* What is probability based on frequency?
* Calculate probability based on belief – Bayes rule

## **From Joint distribution to Graphical models**

* Joint distribution:

Substituting P(a,b) gives

* Converting the above probability to graph:

Node A is parent of Node B and C. Node B and C are children for Node A.

## **Boxes of fruit example**

P(bluebox, apple) = ?

P(bluebox) = 4/12 = 2/6 = 1/3

P(redbox) = 8/12 = 4/6 = 2/3

P(apple| bluebox) = ¾

P(apple| redbox) = 2/8 = ¼

P(orange| bluebox) = ¼

P(orange| redbox) = 6/8 = ¾

P(bluebox,apple) =

## **Joint distribution of K variables**

Repeating the same procedure gives

*Joint probability distribution*



Combining Bayesian and Markov model – Chained model

## **Reasoning patterns:**

* Causal reasoning
  + Relationships between the variables
* Evidential reasoning
  + From effects to causes
* Intercausal reasoning
  + One causal factor gives information about another causal factor
* These concepts are explained in Student model example in the lecture.

## **Learning Bayesian network**

* Known data:
  + Complete data
    - Parameter estimation: Maximum likelihood (ML), Maximum APosteriori (MAP) approach
  + Incomplete data
    - Non-linear parametric estimation (Gradient descent , Expectation Maximization)
* Unknown data
  + Complete data
    - Optimization (search in space of graphs)
  + Incomplete data
    - Structural EM, mixture model

## **Evaluation of Bayesian networks**

* Log likelihood (LL)
* Minimum Description Length

## **Algorithms used for building Bayesian networks**

* K2 algorithm
* TAN (Tree augmented Naïve Bayes)

## **Tricks for Bayesian networks**

* Markov blanket
  + Markov blanket of a node includes all parents, children, and children’s parents of that node
  + Given values for Markov blanket, node is conditionally independent of nodes outside blanket

## **Bayesian multinet**

* Build one network for each class and make prediction using Bayes’ rule