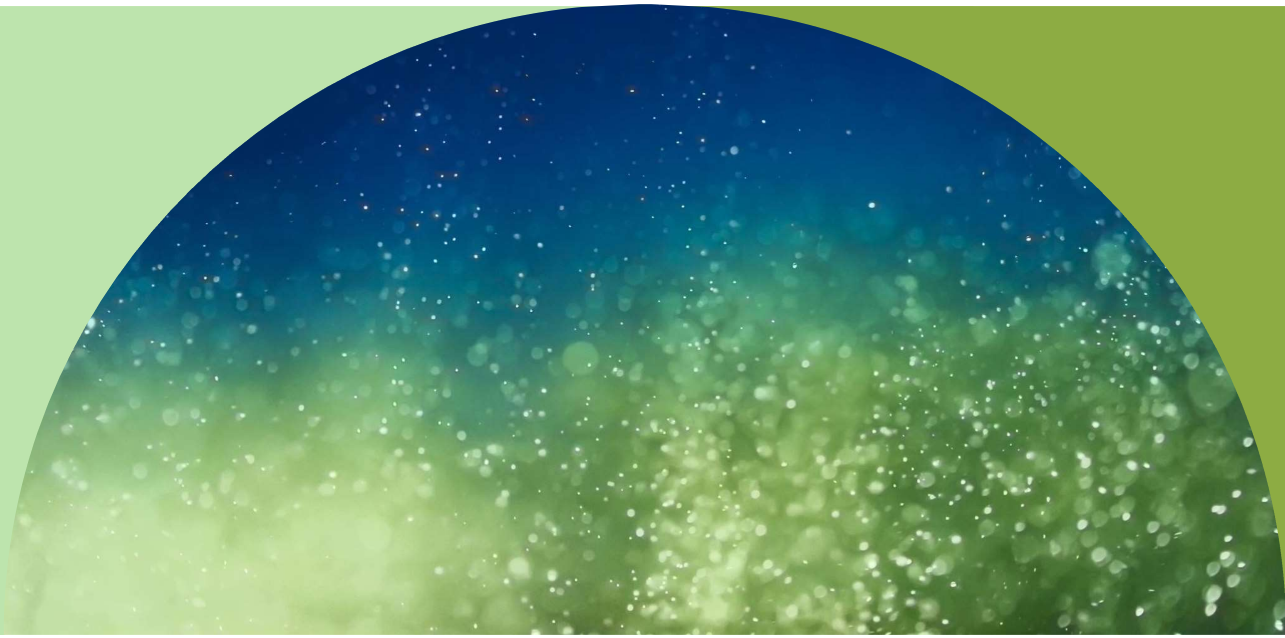
Cognitive Systems : Probabilistic Artificial Intelligence

Rahee Walambe



3	**Formal modeling: Bayesian Inference and Hierarchical Bayesian Models, Frameworks for Knowledge Representation: First-order Logic, Formal Grammars, Associative Networks, Taxonomic Hierarchies, Relational Schemas, Probabilistic and Causal Graphical Models, Relational Probabilistic Models, Controlling Complexity: Minimum Description Length, Bayesian Occam's Razor, Nonparametric Bayesian Models Inductive Logic Programming, Sampling Algorithms for Inference in Complex Probabilistic Models**	**10**	**10**

Topics

1. Probabilistic Computing
2. Third wave of AI
3. Ex: Driving a car
4. Role in Explainable AI (XAI)
5. Role of probability in machine learning

Probability in AI

- Probabilistic computing allows us to

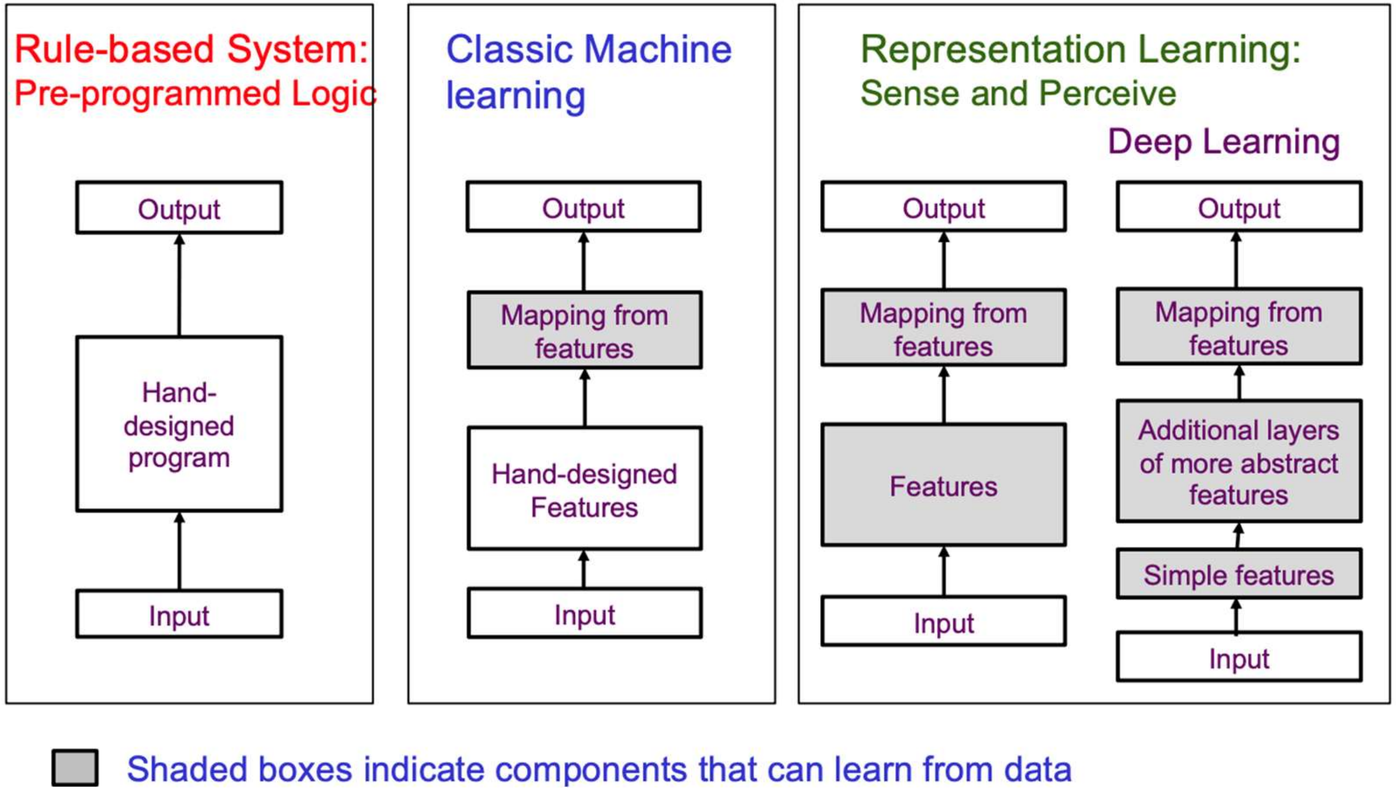
1. Deal with uncertainty in natural data around us
2. Predict events in the world with an understanding of data and model uncertainty

- Predicting what will happen next in a scenario, as well as the effects of our actions, can only be done if we know how to model the world around us with probability distributions
- Research into probabilistic computing is really about establishing a new way to evaluate the performance of the next wave of AI —one that requires real-time assessment of “noisy” data.

Role with XAI

- Augmenting deep learning with probabilistic methods opens the door to understanding why AI systems make the decisions they make,
- Will help with issues like tackling bias in AI systems.

Current AI Models



Next AI models

- First AI systems focused on logic:
- – Pre-programed rules.
- Second wave of AI concerns ability to sense and perceive information
- – Leveraging neural networks to learn over time.
- But, neither solution can do things that human beings do naturally as we navigate the world.
- – They can’t think through multiple potential scenarios based on data that you have on-hand while
- conscious of potential data that you don’t have

Example : Driving a Car and Ball

- If you are driving a car and see a soccer ball roll into the street,
- Your immediate and natural reaction is to stop the car since we can assume a child is running after the ball and isn’t far behind.
- Driver reaches the decision to stop the car based on experience of natural data and assumptions about human behavior.
  - But a traditional computer likely wouldn’t reach the same conclusion in real-time, because today’s systems are not programmed to mine noisy data efficiently and to make decisions based on environmental awareness.
  - You would want a probabilistic system calling the shots—one that could quickly assess the situation and act (stop the car) immediately.

Role of Probability in AI

- In neural networks (discriminative models)

1. Output is a probability distribution over *y*
2. Instead of error as loss function we use a surrogate loss function, viz., log-likelihood, so that it is differentiable (which is necessary for gradient descent)

- In probabilistic AI (generative models)

- We learn a distribution over observed and latent variables whose parameters are determined by gradient descent as well

Overview of Probabilistic Graphical Models

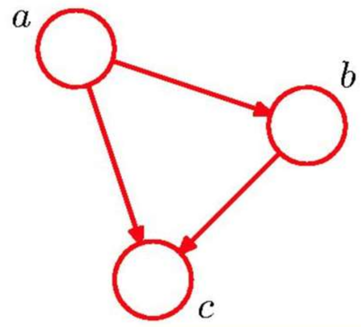
- Probabilistic graphical models (or PGMs)
- Directed and Undirected Graphical Models
- Joint and Conditional Probability Distributions
- Probabilistic Queries and Inference

Probabilistic Graphical models

- They are *diagrammatic representations* of probability distributions
- – marriage between probability theory and graph theory
- Also called *probabilistic graphical models*
- They augment analysis instead of using pure algebra

What is a Graph

- Consists of nodes (also called vertices) and links (also called edges or arcs)
- In a probabilistic graphical model
- – each node represents a random variable (or group of random variables)
- – Links express probabilistic relationships between
- variables



Graphical Models in Engineering

- Natural tool for handling Uncertainty and Complexity
- – which occur throughout applied mathematics and engineering
- Fundamental to the idea of a graphical model is the notion of modularity
- – a complex system is built by combining simpler parts.

Why are Graphical Models useful in Engineering?

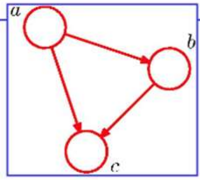
- Probability theory provides the glue whereby
- – the parts are combined, ensuring that the system as a whole is consistent
- – providing ways to interface models to data.
- Graph theoretic side provides:
- – Intuitively appealing interface

  by which humans can model highly-interacting sets of variables
- –Data structure

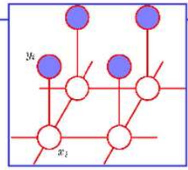
  that lends itself naturally to designing efficient general-purpose algorithms

Graph Directionality

- Directed graphical models
  - directionality associated with arrows
- Bayesian networks (BNs)
  - Express causal relationships between random variables
- More popular in AI and statistics



- Undirected graphical models
  - links without arrows
- Markov random fields (MRFs)
  - Better suited to express soft constraints between variables
- More popular in Vision and physics



Bayesian Networks

- Directed graphs
- – used to describe probability distributions
- Consider Joint distribution
- – of three variables a,b,c
- Powerful aspect of graphical models
- – Not necessary to state whether they are discrete or continuous
- – A specific graph
- • can make probabilistic statements about a broad class of distributions
- Bayesian Network is not necessarily Bayesian statistics

Joint and Conditional Distributions

- The necessary probability theory can be expressed in terms of two simple equations
  - Sum Rule
    - probability of a variable is obtained by marginalizing summing out other variables

$$p(a) = \sum_b p(a,b)$$

  - Product Rule
    - joint probability expressed in terms of conditionals

$$p(a,b) = p(b \mid a)\,p(a)$$

All probabilistic inference and learning amounts to repeated application of sum and product rule

From Joint Distribution to Graphical Model

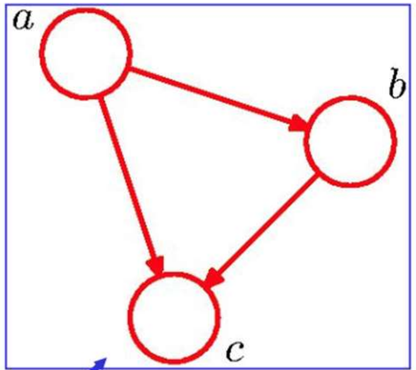
- Consider Joint distribution $p(a,b,c)$
  - By product rule $p(a,b,c)=p(c|a,b)p(a,b)$
  - Again by product rule we get

$$\boxed{p(a,b,c)=p(c|a,b)p(b|a)p(a)}$$

- This decomposition holds
  - for any choice of the joint distribution

Directed Graphical Model

$$p(a,b,c) = p(c|a,b)\,p(b|a)\,p(a)$$

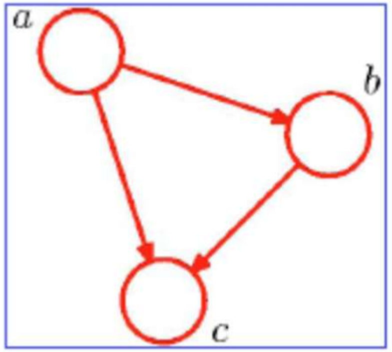


- Now represent rhs by graphical model
  - Introduce a node for each random variable
  - Associate each node with conditional distribution on rhs
    - For each conditional distribution add links (arrow): for $p(c|a,b)$ links from a and b to c
- Different ordering of variables would give a different graph

TERMINOLOGY

- Node a is *parent* of node b
- Node b is *child* of node a
- *No distinction between node and variable*

From Graph to Distribution



$$\boxed{p(a,b,c)=p(c|a,b)p(b|a)p(a)}$$

- Graph says:
  - Distribution is a product of three terms
  - There are three known distributions whose product defines the joint distribution
  - Each distribution could be specified by a table