# Uncertainty in Computation: Artificial Intelligence and Machine Learning

Vasant Honavar
Artificial Intelligence Research Laboratory
College of Information Sciences and Technology
Bioinformatics and Genomics Graduate Program
The Huck Institutes of the Life Sciences
Pennsylvania State University

vhonavar@ist.psu.edu
http://vhonavar.ist.psu.edu
http://faculty.ist.psu.edu/vhonavar

# Representing and Reasoning under uncertainty

- An agent may be uncertain about
  - the state of the world or the results of its actions due to
    - Partial observability
    - Non-determinism
  - its knowledge of the world

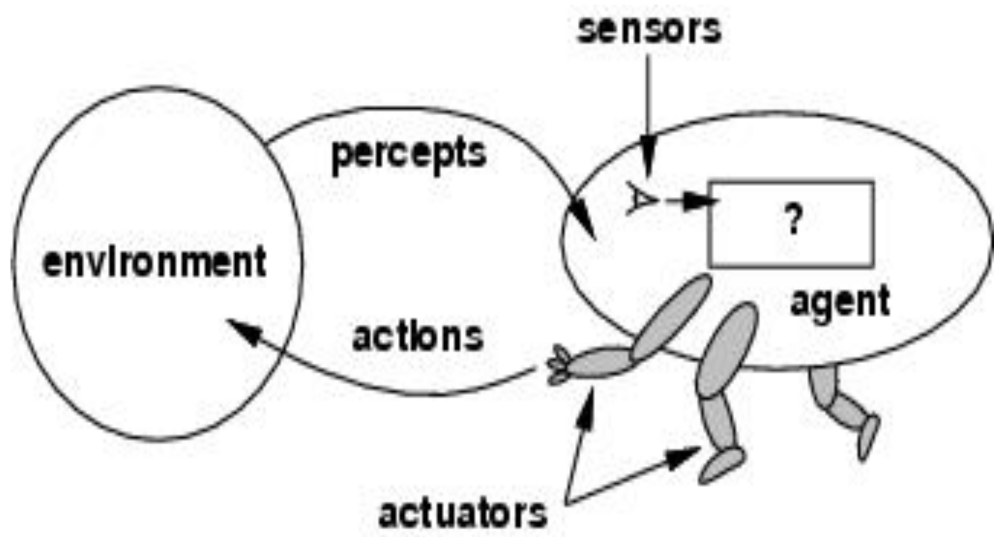



Figure source: Russell and Norvig

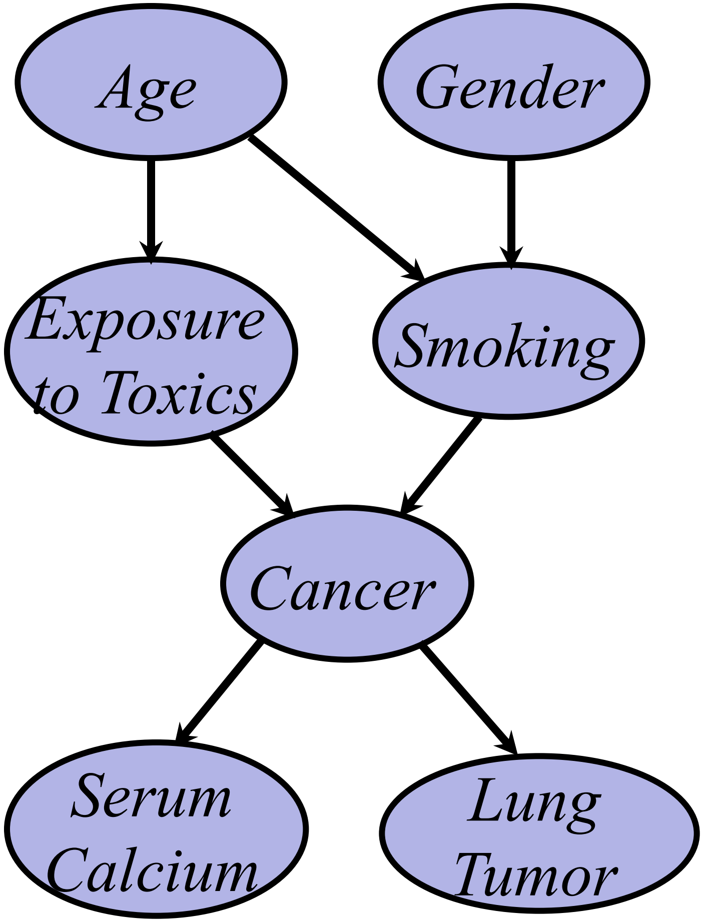
# Uncertainty measures

- Probability measure
- Imprecise probabilities – useful in settings where probabilities are only partially specified, e.g., in terms of upper and lower bounds [Walley, 1991; Kuznetsov, 1991]
- Dempster-Shafer belief functions [Dempster 1967; Shafer, 1976]
- Possibility measures [Zadeh, 1978; Dubois and Prade, 1978]
- Ranking functions [Levi, 1985]
- Plausibility measures – generalization of all of the above [Halpern, 2005]

# Probability Theory as a Knowledge Representation

- Ontological commitments (what can we talk about?)
  - Sentences that represent an agent's beliefs about the world
- Epistemological Commitments (what can we know)
  - The *probability* that a given sentence is true
- Syntax
  - Propositional logic based
- Semantics
  - Possible worlds interpretation $$P(\varphi) = \sum_{\omega \models \varphi} \mu(\omega)$$
- Proof Theory
  - Based on logic and probability

# Propositional probability models – Bayes Networks




Pearl, J. Probabilistic reasoning in intelligent systems. Cambridge Univ. Press. 1998.

- Nodes denote random variables
- DAG encodes a set of conditional independence relationships between random variables
- Each variable is conditionally independent of its non-descendants given its parents

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Parents(X_i))$$

- Factorized representation admits efficient inference

$$P(A,G,E,S,C,L,SC) = P(A)P)(G)P(E \mid A)P(S \mid A,G)P(C \mid E,S)P(SC \mid C)P(L \mid C)$$

# First order probability models

- Can we combine probability with the expressive power of first order logic (FOL) representation?
- Problem: The set of possible worlds represented by an FOL sentence can be infinite
- Relational probability models (RPM) 'solve' this problem by replacing standard FOL semantics by database semantics
  - Unique names assumption (e.g., each customer has a unique ID)
  - Domain closure assumption (there are no more objects beyond the ones that have been named)

Koller, Pfeffer, Getoor et al. 1999-2007

# Open universe probability models

- Unique names assumption and domain closure assumption do not hold in the presence of uncertainty about existence and identity of objects
- Open universe probability models (OUPMs) extend Bayes networks and RPMs by adding
  - generative steps that add objects to the possible world under construction
  - where the number and type of objects added may depend on the objects that are already present

Milch et al., 2007

# Probabilistic Programming Languages

- Logic based
  - PRISM, Problog – logic programming + probability distributions over facts [Sato and Kameya, 2001; De Raedt, Kimmig, and Toivonen, 2007]
  - BLOG – a language based on open universe probability models [Milch et al., 2007]
- Functional programming based
  - Church, Venture – extend Scheme with probabilistic semantics for specifying recursively defined generative processes [Goodman, Mansinghka, Roy, Bonawitz and Tenenbaum, 2008]
  - IBAL – a stochastic functional programming language [Pfeffer, 2007]
- Object-oriented
  - Figaro – an expressive language with support for directed and undirected probabilistic graphical models, OUPMs, models defined over complex data structures. [Pfeffer, 2009]

# Causal Bayesian Networks

- A **causal BN** is a BN with an explicit requirement that the relationships be causal [Pearl, 2000]
- If an intervention on node $X$ (an action written as $do(X=x)$) sets $X=x$ then the distribution changes to that obtained by cutting the links from Parents($X$) to $X$, and setting $X$ to $x$
  - Identifiability – determining whether a given set of causal assumptions is sufficient for determining causal effects from observations and experiments [Tian and Pearl, 2002; Huang, 2006; Shpitser and Pearl, 2006 ..]
  - z-Identifiability – estimating the effect of intervening on a set of variables X from experiments on a different set, Z [Bareinboim and Pearl, 2012]
  - Transportability – reuse of causal information obtained from experiments in one setting in a related but different setting where only observations are available [Bareinboim and Pearl, 2012; Lee and Honavar, 2013; Bareinboim and Pearl 2013]
  - Generalizations of the above [Bareinboim, Lee, Honavar, Pearl, 2013]

# Uncertainty in machine learning

- Data uncertainty
  - missing attribute values, partially specified attribute values
  - missing class labels, partially specified class labels
  - distributional data
- Need to generalize standard algorithms that cope with data uncertainty
  - Partially specified instances [Zhang et al, 2003; 2006]
  - Partially specified class labels e.g., "multiple instance learning", "semi-supervised learning", "structured label learning" …
  - Distributional data [Lin et al., 2013; Muandet et al., 2012]

# Uncertainty in machine learning

- Model uncertainty
  - Is the assumed hypothesis class appropriate? (a form of epistemic uncertainty)
  - Can we trust the model parameters (pre-specified? learned?)
  - Can we trust the reported performance?
  - How can we update uncertainty?
  - How can we propagate uncertainty?