# CSE 150A Notes

## 1 9/26

### 1.1 Course Information

Prerequisites:

- Programming Knowledge
- Elementary Probability
  - Random Variables - discrete and continuous
  - Expected Values - sums and integrals
- Multivariable Calculus
- Linear Algebra

HW Released Tuesdays, due Monday 24 hr late policy for HW Quizzes in person every Thursday lecture – Based on HW Point lost on quizzes go to the Final Midterm: 10/31 (Week 5) in class Final: 12/5 (Week 10) in class One sheet of handwritten notes allowed

### 1.2 Course Overview

- Inference and learning in Bayesian Networks
- Markov decision processes for reinforcement learning

Does not cover:

- Neural architectures
- Purely logical reasoning
- Heuristic search (A*)
- Theorem proving
- Genetic algorithms
- Philosophy of AI

## 2 10/1

Probability Theory: how knowledge affects belief (Poole and Mackworth)
This view is known as the Bayesian view of probability
Other view is the frequentist view; probability is the limit of the relative frequency of an event
Discrete Random Variables: denoted with capital letters
Domain of possible values for a variable, denoted with lowercase letters
Uncoditional (prior) probability: $P(X = x)$
Axioms of Probability:
$P(X = x) \geq 0$
$\sum_{i=1}^{n} P(X = x_i) = 1$
$P(X = x_i \text{ or } X = x_j) = P(X = x_i) + P(X = x_j)$ iff $x_i \neq x_j$
Conditional Probability: $P(X = x_i|Y = y_j)$
In this case X and Y are dependent
Bayes rule: $P(X = x_i|Y = y_j) = \frac{P(Y=y_j|X=x_i)P(X=x_i)}{P(Y=y_j)}$

Product rule: $P(X=x_i, Y=y_j) = P(X=x_i|Y=y_j)P(Y=y_j) = P(x)P(y|x)$
Marginalization: $P(X=x_i) = \sum_{j=1}^{n} P(X=x_i, Y=y_j)$
Independence: $P(X=x_i, Y=y_j) = P(X=x_i)P(Y=y_j)$

# 3 10/3

Marginal Independence: $P(X=x_i, Y=y_j) = P(X=x_i)P(Y=y_j)$
$P(X|Y) = P(X)$
$P(Y|X) = P(Y)$
Conditional Independence: $P(X,Y|E) = P(X|E)P(Y|E)$
$P(X|Y,E) = P(X|E)$
$P(Y|X,E) = P(Y|E)$
Suppose $X_i \in \{0,1\}$; then it requires $O(2^n)$ parameters to represent the joint distribution
Goals: compact representation, efficient inference
Use belief to simplify the joint distribution
Conditional Probability Tables (CPT) represent the conditional probability of a variable given its parents
Any inference can be explained in terms of the joint probability, using product rule and marginalization
To perform inference efficiently:
Visualize models as directed acyclic graphs (DAGs)
Exploit the graph structure to simplify and organize calculations
Absent edges represent assumptions of independence
Visual representation of the joint distribution is called a Bayesian Network or belief network
Nodes represent random variables
Edges represent direct dependencies
CPTs for each node describe how each node depends on its parents
Belief network = DAG + CPTs
It's always true from the product rule that $P(X_1, X_2, ..., X_n) = P(X_1)P(X_2|X_1)\cdots = \Sigma_{i=1}^{n} P(X_i|X_1, X_2, \ldots, X_{i-1})$
But suppose in a particular domain that $P(X_i|X_1, X_2, \ldots, X_{i-1}) = P(X_i|\text{parents}(X_i))$
Where parents is a subset of $X_1, X_2, \ldots, X_{i-1}$

To create a belief network:
Choose random variables of interest, choose ordering of the variables, and:
While there are variables left, add the node $X_i$ to the network, with parents the minimum subset satisfying:
$P(X_1, X_2 \ldots X_n) = \Sigma_{i=1}^{n} P(X_i|\text{parents}(X_i))$
Define CPTs
Best order is to take "root causes", then variables they influence, etc.
Edge does not necessarily represent dependence, especially if a bad order is chosen
DAGs encode qualitative knowledge: assumptions of marginal and conditional independence
CPTs encode quantitative knowledge: numerical influences of some variables on others

# 4 10/8