Artificial Intelligence
Recap & Expectation Maximization
CSE 473

Dan Weld

Exam Topics

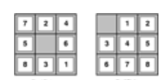
- Search
  - BFS, DFS, UCS, A* (tree and graph)
  - Completeness and Optimality
  - Heuristics: admissibility and consistency
- CSPs
  - Constraint graphs, backtracking search
  - Forward checking, AC3 constraint propagation, ordering heuristics
- Games
  - Minimax, Alpha-beta pruning, Expectimax, Evaluation Functions
- MDPs
  - Bellman equations
  - Value iteration
- Reinforcement Learning
  - Exploration vs Exploitation
  - Model-based vs. model-free
  - Q-learning
  - Linear value function approx.
- Hidden Markov Models
  - Markov chains
  - Forward algorithm
  - Particle Filter
- Bayesian Networks
  - Basic definition, independence
  - Variable elimination
  - Sampling (rejection, importance)
- Learning
  - BN parameters with complete data
  - Search thru space of BN structures
  - Expectation maximization

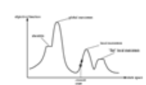
What is intelligence?

- (bounded) Rationality
  - We have a performance measure to optimize
  - Given our state of knowledge
  - Choose optimal action
  - Given limited computational resources
- Human-like intelligence/behavior

Search in Discrete State Spaces

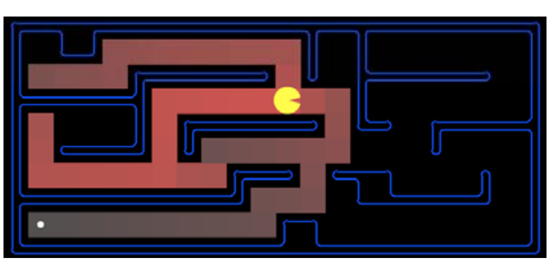
- Every discrete problem can be cast as a search problem.
  - states, actions, transitions, cost, goal-test
- Types
  - uninformed systematic: often slow
    - DFS, BFS, uniform-cost, iterative deepening
  - Heuristic-guided: better
    - Greedy best first, A*
    - relaxation leads to heuristics
  - Local: fast, fewer guarantees; often local optimal
    - Hill climbing and variations
    - Simulated Annealing: global optimal
  - (Local) Beam Search



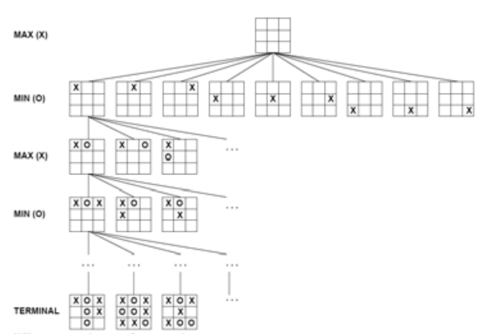


Which Algorithm?

- A*, Manhattan Heuristic:

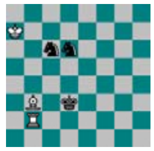


Adversarial Search



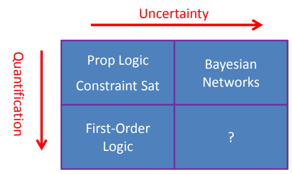
Adversarial Search

- AND/OR search space (max, min)
- minimax objective function
- minimax algorithm (~dfs)
  - alpha-beta pruning
- Utility function for partial search
  - Learning utility functions by playing with itself
- Openings/Endgame databases



Knowledge Representation and Reasoning

- Representing: what I know
- Reasoning: what I can infer



KR&R Example: Propositional Logic

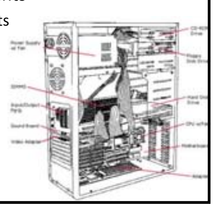
- Representation: Propositional Logic Formula
  - CNF, Horn Clause,…
- Reasoning: Deduction
  - Forward Chaining
  - Resolution
- Model Finding
  - Enumeration
  - SAT Solving

Expressivity

- Propositional Logic vs Bayesian network?
- $(X \wedge Y) \vee (\neg X \wedge \neg Y)$

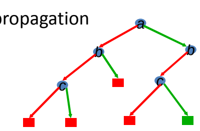
Search+KR&R Example: CSP

- Representation
  - Variables, Domains, Constraints
- Reasoning: Constraint Propagation
  - Node consistency, Arc Consistency, k-Consistency
- Search
  - Backtracking search: partial var assignments
    - Heuristics: min remaining values, min conflicts
  - Local search: complete var assignments



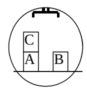
Search+KR&R Example: SAT Solving

- Representation: CNF Formula
- Reasoning
  - pure literals; unit clauses; unit propagation
- Search
  - DPLL (~ backtracking search)
    - MOM's heuristic
  - Local: GSAT, WalkSAT
- Phase Transitions in SAT problems



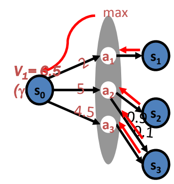
Search+KR&R Example: Planning

- Representation: STRIPS
- Reasoning: Planning Graph
  - Polynomial data structure
  - reasons about constraints on plans (mutual exclusion)
- Search
  - Forward: state space search
    - planning graph based heuristic
  - Backward: subgoal space search
- Planning as SAT: SATPlan



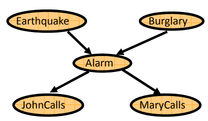
KR&R: Markov Decision Process

- Representation
  - states, actions, probabilistic outcomes, rewards
  - ~AND/OR Graph (sum, max)
  - Generalization of expectimax
- Reasoning: V*(s)
  - Value Iteration: search thru value space
- Reinforcement Learning:
  - Exploration / exploitation
  - Learn model or learn Q-function?



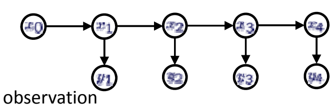
KR&R: Probability

- Representation: Bayesian Networks
  - encode probability distributions compactly
    - by exploiting conditional independences
- Reasoning
  - Exact inference: var elimination
  - Approx inference: sampling based methods
    - rejection sampling, likelihood weighting, MCMC/Gibbs



KR&R: Hidden Markov Models

- Representation
  - Spl form of BN
  - Sequence model
  - One hidden state, one observation
- Reasoning/Search
  - most likely state sequence: Viterbi algorithm
  - marginal prob of one state: forward-backward



Learning Bayes Networks

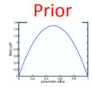
- Learning Structure of Bayesian Networks
  - Search thru space of BN structures
- Learning Parameters for a Bayesian Network
  - Fully observable variables
    - Maximum Likelihood (ML), MAP & Bayesian estimation
    - Example: Naïve Bayes for text classification
  - Hidden variables
    - Expectation Maximization (EM)

Bayesian Learning

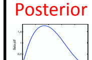
Use Bayes rule:

Data Likelihood

Prior



Posterior

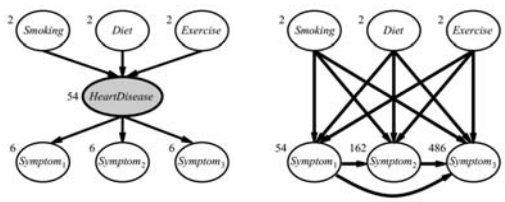


$$P(Y \mid \mathbf{X}) = \frac{P(\mathbf{X} \mid Y)\, P(Y)}{P(\mathbf{X})}$$

Normalization

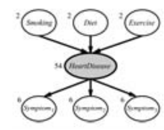
Or equivalently: $P(Y \mid \mathbf{X}) \propto P(\mathbf{X} \mid Y)\, P(Y)$

Why Learn Hidden Variables?



Chicken & Egg Problem

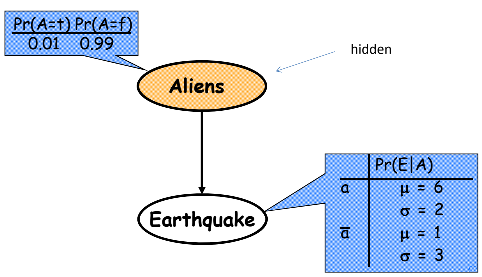
- If we knew whether patient had disease
  - It would be easy to learn CPTs
  - But we can't observe states, so we don't!



- If we knew CPTs
  - It would be easy to predict if patient had disease
  - But we don't, so we can't!

Slide by Daniel S. Weld 20

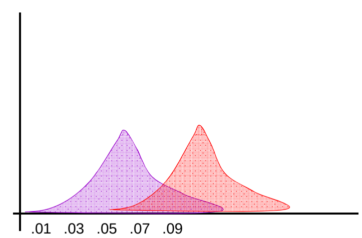
Continuous Variables



© Daniel S. Weld

Simplest Version

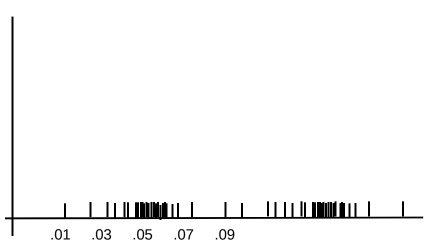
- Mixture of two distributions



- Know: form of distribution & variance,
  % =5
- Just need *mean* of each distribution

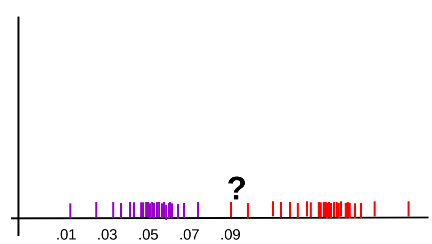
Slide by Daniel S. Weld 22

Input Looks Like



Slide by Daniel S. Weld 23

We Want to Predict

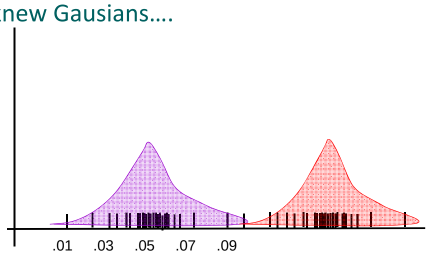


Slide by Daniel S. Weld 24

Chicken & Egg

Note that coloring instances would be easy

if we knew Gausians….



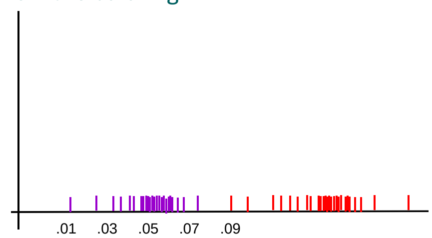
Slide by Daniel S. Weld

25

Chicken & Egg

And finding the Gausians would be easy

If we knew the coloring

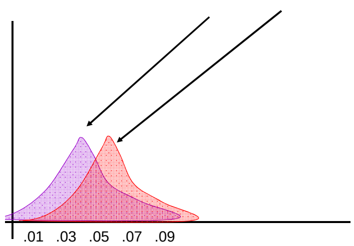


Slide by Daniel S. Weld

26

Expectation Maximization (EM)

- Pretend we *do* know the parameters
  - Initialize randomly: set θ_1=?; θ_2=?

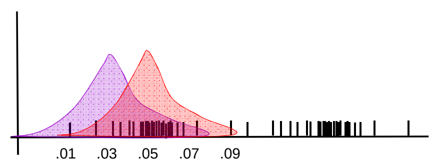


Slide by Daniel S. Weld

27

Expectation Maximization (EM)

- Pretend we *do* know the parameters
  - Initialize randomly
- [E step] Compute probability of instance having each possible value of the hidden variable

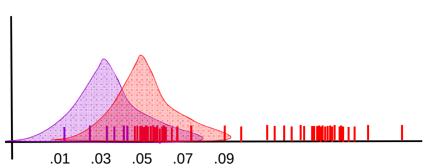


Slide by Daniel S. Weld

28

Expectation Maximization (EM)

- Pretend we *do* know the parameters
  - Initialize randomly
- [E step] Compute probability of instance having each possible value of the hidden variable



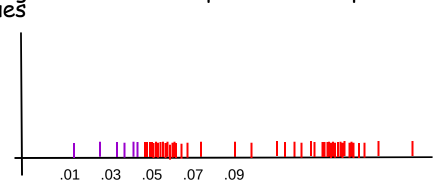
Slide by Daniel S. Weld

29

Expectation Maximization (EM)

- Pretend we *do* know the parameters
  - Initialize randomly
- [E step] Compute probability of instance having each possible value of the hidden variable

[M step] Treating each instance as ***fractionally*** having **both** values compute the new parameter values

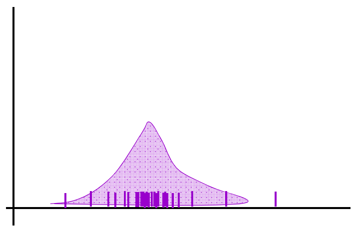


Slide by Daniel S. Weld

30

ML Mean of Single Gaussian

$U_{ml} = \text{argmin}_u \sum_i (x_i - u)^2$

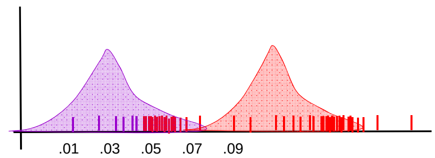


Slide by Daniel S. Weld

31

Expectation Maximization (EM)

[M step] Treating each instance as fractionally having **both** values compute the new parameter values

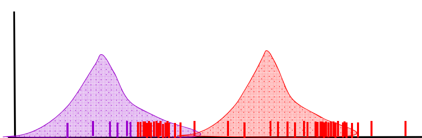


Slide by Daniel S. Weld

32

Expectation Maximization (EM)

- [E step] Compute probability of instance having each possible value of the hidden variable



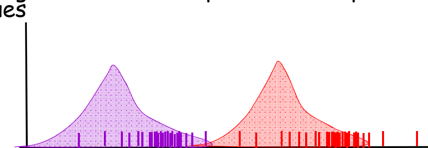
Slide by Daniel S. Weld

33

Expectation Maximization (EM)

- [E step] Compute probability of instance having each possible value of the hidden variable

[M step] Treating each instance as fractionally having both values compute the new parameter values



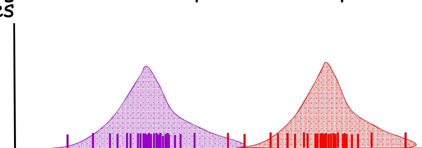
Slide by Daniel S. Weld

34

Expectation Maximization (EM)

- [E step] Compute probability of instance having each possible value of the hidden variable

[M step] Treating each instance as fractionally having both values compute the new parameter values



Slide by Daniel S. Weld

35

Applications of AI

- Mars rover: planning
- Jeopardy: NLP, info retrieval, machine learning
- Puzzles: search, CSP, logic
- Chess: search
- Web search: IR
- Text categorization: machine learning
- Self-driving cars: robotics, prob. reasoning, ML…

Ethics of Artificial Intelligence

- Robots
  - Robot Rights
  - Three Laws of Robotics
- AI replacing people jobs
  - Any different from industrial revolution?
- Ethical use of technology
  - Dynamite vs. Speech understanding
- Privacy concerns
  - Humans/Machines reading freely available data on Web
  - Gmail reading our news
- AI for developing countries/improving humanity

Exam Topics

- Search
  - BFS, DFS, UCS, A* (tree and graph)
  - Completeness and Optimality
  - Heuristics: admissibility and consistency
- CSPs
  - Constraint graphs, backtracking search
  - Forward checking, AC3 constraint propagation, ordering heuristics
- Games
  - Minimax, Alpha-beta pruning, Expectimax, Evaluation Functions
- MDPs
  - Bellman equations
  - Value iteration
- Reinforcement Learning
  - Exploration vs Exploitation
  - Model-based vs. model-free
  - Q-learning
  - Linear value function approx.
- Hidden Markov Models
  - Markov chains
  - Forward algorithm
  - Particle Filter
- Bayesian Networks
  - Basic definition, independence
  - Variable elimination
  - Sampling (rejection, importance)
- Learning
  - BN parameters with complete data
  - Search thru space of BN structures
  - Expectation maximization