CSE 473

Lecture 19
(Chapter 21 & 13)

Q Learning and Uncertainty

© CSE AI faculty + Chris Bishop, Dan Klein, Stuart Russell, Andrew Moore

Today's Outline

- Feature-based Q Learning
- Uncertainty
  - Probability Theory
  - Inference by Enumeration

Recall: Q-Learning

- Online ***sample-based*** Q-value iteration.
- At each time step:
  - Execute action and get new sample (s,a,s',r)
  - Incorporate new sample into **running average of Q**:

$$Q(s,a) \leftarrow (1-\alpha)Q(s,a) + \alpha(r + \gamma \max_{a'} Q(s',a'))$$

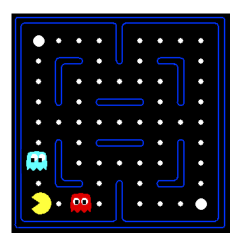
$$\text{where } \alpha \text{ is the learning rate } (0 < \alpha < 1).$$

  - Update policy:

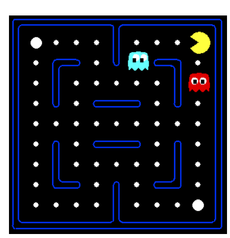
$$\pi(s) = \arg\max_{a} Q(s,a)$$

Problem: Generalization

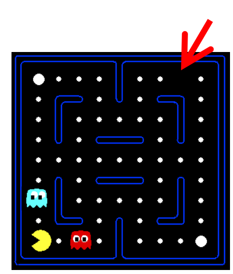
- Let's say we discover through experience that this "trapped" state is bad:



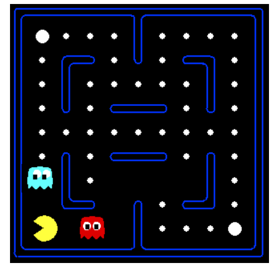
- In naïve Q learning, we know nothing about related states such as this one and their Q values



- Or even this third one!



Feature-Based Representations



- Solution: Describe a state using a vector of features (properties)
  - Features are functions from states to real numbers (often 0/1) that capture important properties of the state
  - Example features:
    - Distance to closest ghost
    - Distance to closest dot
    - Number of ghosts
    - 1 / (dist to dot)^2
    - Is Pacman in a tunnel? (0/1)
    - …… etc.
- Can also describe a Q-state (s, a) with features (e.g. whether action in a state moves closer to food)

Approximating Q-values using Features

- Write a Q function as a linear *weighted combination of feature values*:

$$Q(s,a) = w_1 f_1(s,a) + w_2 f_2(s,a) + \ldots + w_n f_n(s,a)$$

Need to learn the weights w_i – how?

Recall:

We want Q to approximate sample-based average:

$$Q(s,a) \leftarrow \frac{1}{t} \sum_{t \text{ samples}} \left(r + \gamma \max_{a'} Q(s',a') \right)$$

where:

$$Q(s,a) = w_1 f_1(s,a) + w_2 f_2(s,a) + \ldots + w_n f_n(s,a)$$

Find w_i that *minimize error* for each sample:

$$\left| (r + \gamma \max_{a'} Q(s',a')) - Q(s,a) \right|^2$$

Feature-based Q-learning

$$transition = (s, a, r, s')$$

$$\text{Error} = \left[r + \gamma \max_{a'} Q(s',a') \right] - Q(s,a)$$

$$w_i \leftarrow w_i + \alpha [\quad \text{Error} \quad] f_i(s,a)$$

Intuitive interpretation:

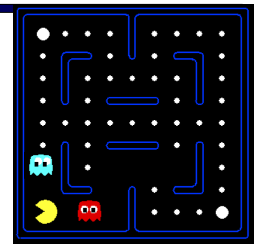
- Weights of active features (f_i is 1 or high value) adjusted
- If a feature is active and the $Q(s,a)$ prediction does not match the desired value: $\left[r + \gamma \max_{a'} Q(s',a') \right]$ then change the weights according to positive/negative error.

Example: Q-Pacman

$$Q(s,a) = 4.0 f_{DOT}(s,a) - 1.0 f_{GST}(s,a)$$

$$f_{DOT}(s, \text{NORTH}) = 0.5$$

$$f_{GST}(s, \text{NORTH}) = 1.0$$



s

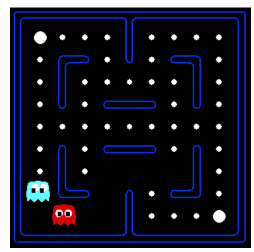
$$Q(s,a) = +1$$

$$R(s,a,s') = -500$$

$a = \text{NORTH}$

$r = -500$

error = -501



s'

$$w_{DOT} \leftarrow 4.0 + \alpha\,[-501]\,0.5$$

$$w_{GST} \leftarrow -1.0 + \alpha\,[-501]\,1.0$$

$$Q(s,a) = 3.0 f_{DOT}(s,a) - 3.0 f_{GST}(s,a)$$

= -1.5

Learning correctly decreases Q value as required!

Q-learning Pac-Man (no features)

Q-learning, no features, 50 learning trials

Q-learning Pac-Man (no features)

Q-learning, no features, 1000 learning trials

Q-learning Pac-Man (with features)

Feature-based Q-learning, 50 learning trials

What if Pac-Man does not know the exact state and only gets local sensor readings about the state
(e.g., camera, laser range finder)?

Enter Uncertainty…

Example: Catching a flight

- Suppose you have a flight at 6pm
- When should you leave for SeaTac?
  - What are the traffic conditions?
  - How crowded is security?

Leaving time before 6pm	P(arrive-in-time)
20 min	0.05
30 min	0.25
45 min	0.50
60 min	0.75
120 min	0.98
1 day	0.99999

Probability Theory: Beliefs about events
Utility theory: Representation of preferences

Decision about when to leave depends on both:
Decision Theory = Probability + Utility Theory

What Is Probability?

- Probability: Calculus for dealing with nondeterminism and uncertainty
- Where do the numbers for probabilities come from?
  - Frequentist view (numbers from experiments)
  - Objectivist view (numbers inherent properties of universe)
  - Subjectivist view (numbers denote agent's beliefs)

Why Should You Care?

- The world is full of uncertainty
  - Incomplete knowledge of the world
  - Noisy sensor readings
  - Ambiguous sensor readings (e.g., images)
- Probability: new foundation for AI (& CS!)
- "Big Data" is today's buzz word!
  - Statistics and CS are both about data
  - Statistics lets us summarize and understand it
  - Statistics is the basis for most learning

Statistics + CS =
Hope + Change

(Nate Silver)

Logic *vs.* Probability

Symbol: Q, R, …	**Random variable**: Q, R, …
Boolean values: T, F	**Values/Domain**: you specify e.g. {heads, tails}, Reals
State of the world: Assignment of T/F to all symbols Q, R …	**Atomic event**: a complete assignment of values to Q, R, … • Mutually exclusive • Exhaustive

Types of Random Variables

Propositional or Boolean random variables
    e.g., $Cavity$ (do I have a cavity?)

Discrete random variables (*finite* or *infinite*)
    e.g., $Weather$ is one of $\langle sunny, rain, cloudy, snow \rangle$
    $Weather = rain$ is a proposition
    Values must be exhaustive and mutually exclusive

Continuous random variables (*bounded* or *unbounded*)
    e.g., $Temp = 21.6$; also allow, e.g., $Temp < 22.0$.

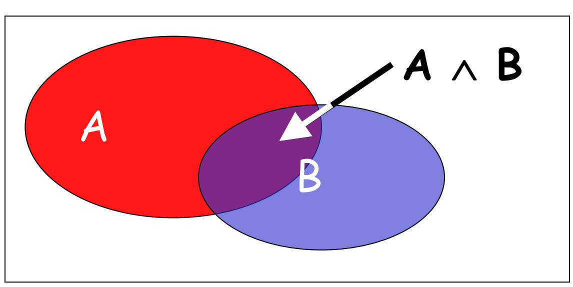
Arbitrary Boolean combinations of basic propositions

Axioms of Probability Theory

Just 3 are enough to build entire theory!

1. All probabilities between 0 and 1
    $0 \leq P(A) \leq 1$
2. P(true) = 1 and P(false) = 0
3. Probability of disjunction of events is:

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$



Prior Probability

Prior or unconditional probabilities of propositions

e.g., $P(Cavity = true) = 0.2$ and $P(Weather = sunny) = 0.72$

correspond to belief prior to arrival of any (new) evidence

Probability distribution gives values for all possible assignments:

$\mathbf{P}(Weather) = \langle 0.72, 0.1, 0.08, 0.1 \rangle$ (*normalized*, i.e., sums to 1)

sunny, rain, cloudy, snow

Joint Probability

Joint probability distribution for a set of r.v.s gives the probability of every atomic event on those r.v.s

$\mathbf{P}(Weather, Cavity) =$ a 4×2 matrix of values:

$Weather =$	$sunny$	$rain$	$cloudy$	$snow$
$Cavity = true$	0.144	0.02	0.016	0.02
$Cavity = false$	0.576	0.08	0.064	0.08

Next time, we will see how any question can be answered by the joint distribution

Next Time

- Probabilistic Inference
- Conditional Independence
- Bayesian Networks
- To Do
  - Project 3
  - Chapter 13 and 14