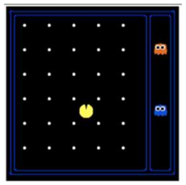
- Reflex: select action based on present. No Memory.
- Model Based: Select action based on past & present
- Goal Based: past, present, future considered
- Utility Based: Past, present, future. Assigns utility score to states
- Learning: Agent can predict. Learns best action through feedback and examples
- Agent: an entity that perceives and acts. Perceives via sensors (percepts), sight sound and taste. Acts via actuators (actions), walk and talk
- Percept: The agent's perceptual inputs at any given instant. Percept Sequence: The complete history of everything the agent has ever perceived.
- Agent function: Maps a percept sequence to an action. External characterization is the function. Implemented by the agent program
- Rational Agent: One that does the "right" thing. Should select an action that is expected to maximize its performance measure.
  - Based on four principles: Performance measure (utility), environment familiarity, actions that are possible, sequences of percepts (memory)
- PEAS: Performance, Environment, Actuators, Sensors
- Task Environments: Deterministic vs Stochastic
  - Deterministic: No randomness when you move a piece. Effect of moving a piece is completely predetermined
  - Stochastic: Can't predict the outcome of the dice. There is randomness present in the environment
- Task Environments: Discrete vs Continuous:
  - Discrete: Finitely many action choices
  - Continuous: Infinitely many action choices
- Task Environments: Benign vs Adversarial
  - Benign: Environment might be stochastic, but there is no objective that would contradict your own objective
  - Adversarial: An opponent is out to get you (like in games)
- Task Environments: Episodic vs Sequential:
  - Episodic: Agents divided into atomic episodes. In each episode, the agent receives a percept and then performs a single action. Next episode is not dependent on the actions taken in previous episode (classification tasks)
  - Sequential: Current decision could affect all future decisions.
- Task Environments: Static vs Dynamic:
  - Static: The environment remains constant
  - Dynamic: The environment can change while the agent is deliberating
- Task Environments: Fully observable/Partially Observable/Unobservable:
  - Fully Observable: Agent can observe all relevant aspects of the environment, no need to maintain any internal state to keep track of the world.
  - Partially Observable: Some parts missing from sensor data. Need memory on the part of the agent to make the optimal decision
  - Unobservable: Can't make any observations but still make a choice.

- Sample/state space: The set of all possible world configurations, or states. Example: Robot vacuum.
- State space graph: A mathematical representation of the problem. Each state is a vertex on the graph, directed edges connect states by corresponding agent actions.

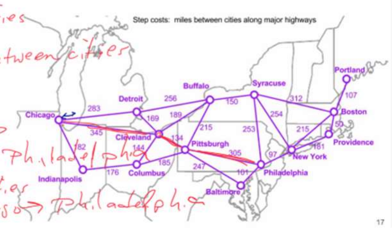


- Size of state space:
  - Location of Pacman: 10 * 12 = 120
  - Pacmans orientation = 4 (u,d,l,r)
  - Food config: 2^{30} Pellet is either eaten or not and there are 30 of them on board
  - Position of ghosts 144. Two ghosts exist in the cage, there are 12 spaces that moves
  - Multiply the values to get the state space (120 * 4 * 2^{30} * 144)
- Search Problem:
  - State Space: What are all the possible ways the world could look, sum of the step costs
  - Transition model: Successor – any state reachable from a given state by a single action
  - Actions: What can the agent do
  - Initial State and Goal test: Determines the start point and whether a given state is the goal state

Search

Example: Traveling in the US northeast Chicago → Philadelphia

1. State space cities
2. Transition model driving between cities
3. Actions driving
4. Initial state Chicago
5. Goal test Are we in Philadelphia
6. Solution Sequence of cities from Chicago → Philadelphia



17

- Markov Decision Process (MDP):
  - Type of search problem, sequential decisions, fully observable, stochastic action/environment
  - Markovian transition model – next state depends only on current, not the past
  - Additive reward structure – maximize reward
  - Used frequently for: Inventory management, logistics, games
  - Requires: states and actions
  - Transition Model (P(s' | s, a)): Probability of being in state s' given action a in states. Multiple successor state from s given action a
  - Reward function (R(s)) or (R(s, a, s'))
  - Policies: Tell the agent what to do in each possible state s, so that the agent can reach the goal (treasure)
  - Rewards: short term gain
  - Utility: long term gain. Sum of rewards along the sequence of states
  - Time Horizon: Finite Horizon after fixing some time N, nothing matters, length of horizon could affect optimal move for a given state. Infinite Horizon: No reason to behave differently in the same state at different times.

  - Discounting: Preference for immediate reward as opposed to future rewards.
  - Optimal Policy: Has highest expected utility

→ Optimal policy in s is then: action

$$\pi^*(s) = \arg\max_{a\in A(s)} \sum_{s'} P(s' \mid s, a)\, U(s')$$

- Value Iteration: Keeps track of utilities and terminates when updates < threshold. to find the optimal utilities.

Start with some candidate set of utilities for each state. Do the following many times (faster than policy iteration, terminates algorithm when no policy changes, fewer iterations than value iteration and actions in all states are unchanged):

  - For each state s: What are the actions available? For each action a: what are the next states, and with what

$$U_{i+1}(s) = \underbrace{R(s)} + \gamma \underbrace{\max_{a\in A(s)} \sum_{s'} P(s'|s,a)\, \overset{0}{U_i(s')}}$$

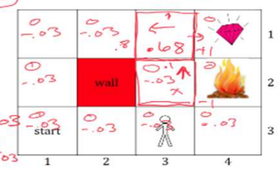
R(s): .03 $U_0 = \phi$ $\gamma = .9$

Example: Find U_0 and U_1 for all states. ϕ

$U_0 = \phi$ for all states
$U_1 = R(s)$ for all states
$U_2(1,3) = -.03 + .9$

→ .8 x 1 + .1 x -.03 + .1 x -.03
↑ .8 x -.03 + .1 x -.03 + .1 x 1
← .8 x -.03 + .1 x -.03 + .1 x -.03
↓ .8 x -.03 + .1 x -.03 + .1 x +1

$U_2(1,3) = -.03 + .9 \,\underset{\text{max}}{[}.794, .073, -.03, .073] = .68$



probabilities? Calculate expected utility w/ action a. Update utility of s to max of discounted expected utilities, plus reward of s

- Policy Iteration: Stores policies and terminates when no policy changes

Policy Iteration

Example: Given the following grid, find the optimal policy of each state using the policy iteration algorithm. The PolicyEvaluation() function sets U for all states using the current policy, U, and the MDP. Use policy iteration algorithm on slide 5

- The terminal states are a and e, and those states have the rewards shown. Let $\gamma = 0.9$
  $U'(s) = R(s) + \gamma \max [P(s'|s,a)\, U(s)]$
- The actions are move left and move right.
- $P(s'|a, s) = 0.80$ success and 0.20 that the agent stays in the same state.
  $\pi = \rightarrow$, $U(s) = \phi$
- $R(s) = -0.04$

After 1 iteration

	→	→	→	
10	-.04	-.04	-.04	1
a	b	c	d	e

Policy Iteration – This is the third iteration. If you do another iteration you'll find that the policy for state d does change to move left.

Example: (continued)

next iteration

$U(b) = -.04 + \gamma \max \begin{cases} \leftarrow .8 \times 10 + .2 \times 7.15 \\ \rightarrow .8 \times -.076 + .2 \times 7.15 \end{cases}$ policy unchanged

$= -.04 + .9(.8 \times 10 + .2 \times 7.15) = 8.44$

$U(c) = -.04 + .9 \max \begin{cases} \leftarrow .8 \times 7.15 + .2 \times -.076 \\ \rightarrow .8 \times .672 + .2 \times -.76 \end{cases}$ policy updated to left

$= -.04 + .9(.8 \times 7.15 + .2 \times -.076) = 5.09$

$U(d) = -.04 + .9 \max \begin{cases} \leftarrow .8 \times -.076 + .2 \times .672 \\ \rightarrow .8 \times 1 + .2 \times .672 \end{cases}$ policy unchanged

$= -.04 + .9(.8 \times 1 + .2 \times .672) = .8$

10	← 8.44	← 5.09	→ .8	1
a	b	c	d	e

- Conditional Probability of V = v given C = c, p(V|C), is the probability that the card value is v if you know that the card color is c. This is given by p(V|C) = p(V,C)/p(C)
  - Example: p(A|B) is the probability of event A occurring, given that event B occurs. For example, given that you drew a red card, what's the probability that it's a four (p(four|red))=2/26=1/13. So out of the 26 red cards (given a red card), there are two fours so 2/26=1/13.
- Marginal Probability: the probability of an event occurring (p(A)) in isolation. It may be thought of as an unconditional probability. It is not conditioned on another event. Example: the probability that a card drawn is red (p(red) = 0.5). Another example: the probability that a card drawn is a 4 (p(four)=1/13).
- Joint probability: p(A ∩B). Joint probability is that of event A and event B occurring. It is the probability of the intersection of two or more events. The probability of the intersection of A and B may be written p(A ∩ B). Example: the probability that a card is a four and red =p(four and red) = 2/52=1/26. (There are two red fours in a deck of 52, the 4 of hearts and the 4 of diamonds).
- Bayes Rule: p(A | B) = p(B | A) p(A)/ p(B)
- Evidence, (p(x)): the probability of encountering the data x, independent of class labels.
  - How likely is it that we would, receive an email containing the words x?
- We could calculate p(x) using the Law of total probability

$$p(x) = \sum_c p(x \mid y = c)\, p(y = c)$$

- HMM:
  - Filtering: Type of inference. Compute the belief state given the evidence observed so far ($P(X_t \mid e_{1:t})$)
  - Prediction: type of inference. Compute the distribution over the future state given all evidence up to current state ($P_{t+k} \mid e_{1:t}$), where k > 0 (typically use k = 1)
  - Smoothing: compute the belief state given the evidence over a previous state using all evidence up to current state ($P(X_k \mid e_{1:t})$), where 0 <=k < t

EXAMPLES FROM HOMEWORK:

- Probability:

3.0.4 a. What is the probability of drawing a Jack of any suit?

The probability of drawing a jack of any suit would be the total number of jacks divided by the total number of cards resulting in the following:

$$\frac{4}{52} = \frac{1}{13}$$

3.0.5 b. What is the probability of drawing a card that is of a black suit?

The probability of drawing a card that is of a black suit would be the total number of black cards divided by the total number of cards resulting in the following:

$$\frac{26}{52} = \frac{1}{2}$$

3.0.6 c. What is the probability of drawing a Jack OR a card of a black suit?

The probability of drawing a jack or a card of black suit would the probability of drawing a jack added to the probability of drawing a card that is of a black suit that is NOT a jack, resulting in the following:

$$\frac{4}{52} + \frac{24}{52} = \frac{28}{52}$$

3.0.7 d. Describe in your own words what a joint probability is.

A joint probabiliy is a chance whether two events can happen together and/or at the same time.

3.0.8 2. Rolling Dice

Consider two die with 6 equally weighted sides.

3.0.9 a. What is the probability of a dice roll summing to 7?

There are six possible combinations of the two die to roll for a total of 7. They are the following: (1 and 6), (5 and 2), (4 and 3), (6 and 1), (2 and 5), (3 and 4). Knowing this we can add them all up to get the probability of the dice roll.

$$\frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36}$$

3.0.10 b. What is the probability of a dice roll summing to an even number 10 times in a row?

The chance of a dice rolling to an even number is $\frac{1}{2}$. In order for the sum to be even 10 times in a row, we can find the probability by the following:

$$(\frac{1}{2})^{10} = \frac{1}{1024}$$

3.0.11 c. Let's say that you roll one dice at a time. Your first die comes up as an even number. What is the probability that your dice will sum to 7 after you roll the second die?

Each value has a pair that will be able to be equivalent to 7, making each chance of summing 7 from two die:

$$\frac{1}{6}$$

	True_Fashion	True_Unfashion
Pred_Fashion	197	22
Pred_Unfashion	112	1491
Total	309	1513

(a) If a photo is actually fashionable, what is the chance the ML classifier correctly identified the photo as being fashionable?

The chance the ML classifier correctly identified the photo as being fashionable can be found applying the following:

$$\frac{TfandPf}{Tf}$$

$$\frac{197}{309}$$

(b) We sample a photo from the data set and learn the ML algorithm predicted this photo was not fashionable. What is the probability that it was incorrect and the photo is actually fashionable?

The probability that it was incorrect and the photo is actually fashionable can be found with the following:

$$\frac{112}{112 + 1491}$$

(c) What is the difference between a joint probability and a marginal probability?

Joint probability is a probability where two events will occur together at the same time, whereas a marginal probability operates outside of any variable independently, not influenced by any other variables.

(d) What is the probability that the machine learning system predicted that an item of clothing was fashionable? Is this a joint or a marginal probability?

The probability that the machine learning system predicted that an item of clothing was fashionable could be found with the following:

$$\frac{219}{1822} = 0.1201975851$$

Two balls are placed in a box as follows: A fair coin is tossed and a white ball is placed in the box if a heads occurs, otherwise a red ball is placed in the box. The coin is tossed again and a red ball is placed in the box if a tails occurs, otherwise a white ball is placed in the box. Balls are drawn from the box three times in succession (always replacing the drawn ball back in the box). It is found that on all three occasions a red ball is drawn. What is the probability that both balls in the box are red?

The probability that both balls in the box are red can be found with the following equation:

$$P(RB|R3) = \frac{P(R3|RB) * P(RB)}{P(R3)}$$

Since it is found that on all three occasions that the ball drawn is red, each with a 50% chance of actually being able to draw a red ball, we can denote the following:

$$P(R3|RB) = 1$$
$$P(RB) = \frac{1}{4}$$
$$P(R3) = (\frac{1}{4} * 1) + (\frac{1}{4} * \frac{1}{8}) + (\frac{1}{4} * \frac{1}{8}) + (\frac{1}{4} * 0)$$
$$P(R3) = \frac{5}{16}$$

Resulting in:

$$P(RB|R3) = \frac{1 * \frac{1}{4}}{\frac{5}{16}}$$
$$P(RB|R3) = \frac{4}{5}$$

A secret government agency has developed a scanner that determines whether a person is an alien imposter. The scanner detects alien imposters with 95% reliability. However, 1% of the time the test returns a false positive for an upstanding citizen. The alien imposters are very dangerous, but also quite rare, with a prevalence in the population of only 0.1%. Upon a tip from an informant, the agency has administered the scanner to a suspicious person.

i. Use the *Law of Total Probability* to calculate the probability that the suspicious person will test positive.

ii. What is the probability that the person is actually an alien imposter given that they do test positive?

To calculate the probability that the suspicious person will test positive, we can denote the following:

$$P(PI) = (0.999 * 0.01) + (0.001 * 0.95)$$
$$P(PI) = 0.0109$$

Or, 1.09%

The probability that the person is actually an alien imposter given that they test positive can be found with:

$$P(PI|I) = 0.95$$
$$P(FI|NI) = 0.01$$
$$P(I) = 0.001$$
$$P(PI) = (0.999 * 0.01) + (0.001 * 0.95)$$
$$P(PI) = 0.0109$$
$$P(I|PI) = \frac{0.95 * 0.001}{0.0109}$$
$$P(I|PI) = 0.087$$

Or, 8.7%

A defective lamp is returned to the company. What is the probability the lamp was produced in factory A?

In order to find the the chance that the lamp was produced in factory A, we need to first find the total defective rate of the lamp in all the factories in the first place:

$$P(D) = (0.35 \cdot 0.015) + (0.35 \cdot 0.010) + (0.30 \cdot 0.020)$$
$$= 0.01475$$

We then divide the total defective rate with the percentage of the total production times the probability of defective lamps for Factory A to yield our final result. This is utilizing Bayes Theorem.

$$= \frac{P(D|A) \cdot P(A)}{P(D)}$$
$$= \frac{.35 \cdot .015}{0.01475}$$
$$= 0.3559322$$

Problem 6 - More likely they have a cold.
Sequence: H → C → C

a. Initial: Healthy = 1.0, Cold = 0.0

Transitions

From	To	
H	H	5/7
H	C	2/7
C	H	1/5
C	C	4/5

Emissions

Evidence	State	
Good	H	6/9 = 2/3
Tired	H	3/9 = 1/3
Good	C	1/6
tired	C	5/6

Day 1

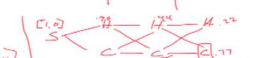
$P(x_1) = \sum_{x_0} P(x_1|x_0)P(x_0)$
$= [5/7 \times 1 + \frac{1}{5} \times 0, \frac{4}{5} \times 0 + 2/7 \times 1]$
$= (5/7, 2/7)$

$P(x_1|\text{good}) = \alpha P(\text{good}|x_1)P(x_1)$
$= \alpha[2/3 \times \frac{5}{7}, \frac{1}{6} \times 2/7]$
$= \alpha[\underset{H}{.47}, \underset{C}{.048}]$

$\alpha = \frac{1}{.47 + .048} = .5176$
$= [\underset{H}{.90}, \underset{C}{.09}]$

Day 2

$P(x_2) = \sum_{x_1} P(x_2|x_1)P(x_1)$
$= [5/7 \times .90 + \frac{1}{5} \times .09, \frac{4}{5} \times .09 + \frac{2}{7} \times .90]$
$= [.66, .329]$



$P(x_2|\text{tired}) = \alpha P(\text{tired}|x_2)P(x_2)$
$= \alpha[1/3 \times .66, 5/6 \times .329]$
$= \alpha[.2178, .274] = [.44, .55]$

Day 3

$P(x_3) = \sum_{x_2} P(x_3|x_2)P(x_2)$
$= [5/7 \times .44 + 1/5 \times .55, 4/5 \times .55 + 2/7 \times .44]$
$= [.424, .565]$

$P(x_3|\text{tired}) = \alpha P(\text{tired}|x_3)P(x_3)$
$= \alpha[1/3 \times .424, 5/6 \times .565]$
$= \alpha[.139, .47]$
$= [.22, .77]$

Day	Person 1	Feeling	Person 2	Feeling	Person 3	Feeling
Day 1	Healthy	Good	Healthy	Good	Healthy	Good
---	---	---	---	---	---	---
Day 2	Cold	Tired	Healthy	Good	Healthy	Good
---	---	---	---	---	---	---
Day 3	Cold	Tired	Healthy	Tired	Cold	Good
---	---	---	---	---	---	---
Day 4	Cold	Tired	Healthy	Tired	Cold	Tired
---	---	---	---	---	---	---
Day 5 (current)	Healthy	Tired	Healthy	Good	Cold	Tired