687 2017-09-26

Note Title 9/26/2017

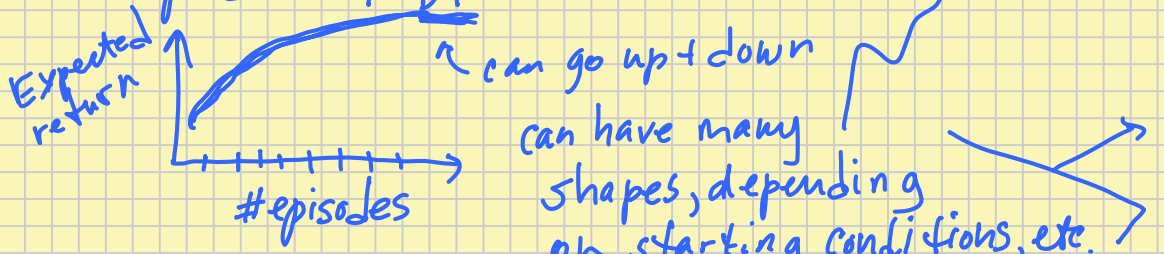
Possible later projects

1) Back pain
2) Car safety

Learning curves

Plot from CEM optimization

Any agent interacting w/ an episodic MDP



can go up + down

can have many shapes, depending on starting conditions, etc.

so:

- Run experiment a number of times, plot average (as estimate of expected reward)
- Plot standard deviation or standard error around the average

Partially Observable Markov Decision Process (POMDP)

$M = (\underline{S}, \underline{a}, \underline{\Omega}, \underline{O}, \underline{P}, \underline{R}, \underline{d_0}, \underline{\gamma})$ _ = same as in an MDP

$\underline{\Omega}$: set of possible observations

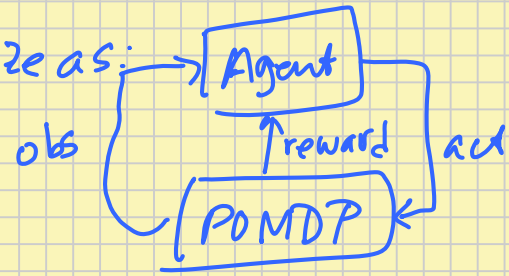
$\underline{O}$: observation function

$O: S \times \Omega \Rightarrow [0,1]$

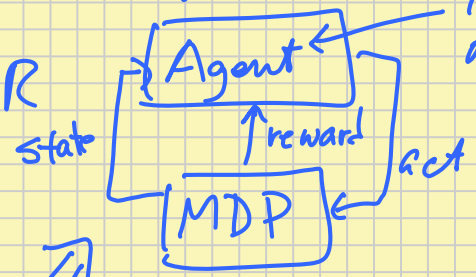
$O(s,\omega) = Pr(W_t = \omega \mid S_t = s)$

W_t: observation at time t

Can organize as:



OR



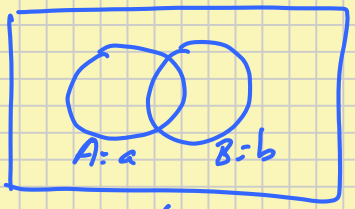
implement partial observability in the Agent

What this course uses

Review of probability:

Conditional probability: $Pr(A=a|B=b)$

$E[A|B=b] = \sum_a a \, Pr(A=a|B=b)$



$\rightarrow \frac{Pr(A=a, B=b)}{Pr(B=b)}$, means "and"

Marginalization:

A = sum of two die rolls

B = first die roll

$Pr(A=a) = \sum_b Pr(B=b) Pr(A=a|B=b)$

Bayes Theorem:

$Pr(A=a|B=b) = \frac{Pr(A=a, B=b)}{Pr(B=b)} =$

$Pr(B=b|A=a) = \frac{Pr(A=a, B=b)}{Pr(A=a)}$

$= \frac{Pr(A=a, B=b) \, Pr(B=b)}{Pr(A=a) \, Pr(B=b)}$

$= \frac{Pr(A=a|B=b) \, Pr(B=b)}{Pr(A=a)}$

Useful for computing back through chains of state transitions, etc.

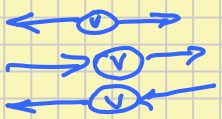
Conditional Independence:

If X is independent of Z given Y

then $Pr(X=x|Y=y, Z=z) = Pr(X=x|Y=y)$

d-separation: X and Z are d-separated by evidence variables Y iff every undirected path from X to Z is blocked.

$\exists v \in Y$ s.t.



d-separated nodes are conditionally independent.

$$Pr(S_2 = s'') = \sum_s Pr(S_0 = s)\, Pr(S_2 = s'' \mid S_0 = s) \quad \text{marginalize over } s$$

$$= \sum_s Pr(S_0 = s) \sum_a Pr(A_0 = a \mid S_0 = s)\, Pr(S_2 = s'' \mid S_0 = s, A_0 = a)$$

conditional indep.

$$= \sum_s Pr(S_0 = s) \sum_a Pr(A_0 = a \mid S_1 = s) \sum_{s'} Pr(S_1 = s' \mid S_0 = s, A_0 = a)\, Pr(S_2 = s'' \mid \cancel{S_0 = s}, \cancel{A_0 = a}, S_1 = s')$$

Our definitions → $d_0(s)$ $\pi(s,a)$ $P(s,a,s')$

$$= \sum_s d_0(s) \sum_a \pi(s,a) \sum_{s'} P(s,a,s') \sum_{a'} \pi(s',a')\, P(s',a',s'')$$

$$Pr(S_2 = s'' \mid S_1 = s', S_0 = s) = Pr(S_2 = s'' \mid S_1 = s') = \sum_{a'} \pi(s',a')\, P(s',a',s'')$$

$$E[R_0] = \sum_s d_0(s) \sum_a \pi(s,a) \sum_{s'} P(s,a,s')\, R(s,a,s')$$

↳ $E[R_0 \mid S_0 = s, A_0 = a, S_1 = s']$

Bellman Equation again: $v^\pi(s) \doteq E\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k} \mid S_t = s, \pi\right]$

$$q^\pi(s,a) \doteq E\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k} \mid S_t = s, A_t = a, \pi\right]$$

$$v^\pi(s) = E\left[R_t + \sum_{k=1}^{\infty} \gamma^k R_{t+k} \mid S_t = s, \pi\right] = E\left[R_t + \gamma \sum_{k=0}^{\infty} \gamma^k R_{t+1+k} \mid S_t = s, \pi\right]$$

$$= \sum_a \underbrace{Pr(A_t = a \mid S_t = s, \pi)}_{\hookrightarrow \pi(s,a)} E\left[R_t + \gamma \sum_{k=0}^{\infty} R_{t+1+k} \mid S_t = s, A_t = s, \pi\right]$$

$$= \sum_a \pi(s,a) \sum_{s'} \frac{P_r(S_{t+1}=s' \mid S_t=s, A_t=a, \pi)}{P(s,a,s')} E\left[R_t + \gamma \sum_{k=0}^{\infty} \gamma^k R_{t+1+k} \mid S_t=s, A_t=a, S_{t+1}=s', \pi\right]$$

$$= \sum_a \pi(s,a) \sum_{s'} P(s,a,s') \left[R(s,a,s') + \gamma E\left[\sum_{k=0}^{\infty} \gamma^k R_{t+1+k} \mid \cancel{S_t=s}, \cancel{A_t=a}, S_{t+1}=s', \pi\right]\right]$$

cannot affect R_{t+1}, R_{t+2}, etc.

$$= \sum_a \pi(s,a) \sum_{s'} P(s,a,s') \left[R(s,a,s') + \gamma v^{\pi}(s')\right]$$

$$\hookrightarrow v^{\pi}(s) = \sum_a \pi(s,a) \sum_{s'} P(s,a,s') \left[R(s,a,s') + \gamma v^{\pi}(s')\right]$$

Can solve as a set of simultaneous equations if $\mathcal{S}$ is finite, etc.

Can develop a similar equation for q.

$$\hookrightarrow q^{\pi}(s,a) = \sum_{s'} P(s,a,s') \left[R(s,a,s') + \gamma v^{\pi}(s')\right]$$

But $v^{\pi}(s') = \sum_{a'} q^{\pi}(s',a')$

$$q^{\pi}(s,a) = \sum_{s'} P(s,a,s') \left[R(s,a,s') + \gamma \sum_{a'} q^{\pi}(s',a')\right]$$

Optimal Value Function $v^*: \mathcal{S} \to \mathbb{R}$

$$v^*(s) = \max_{\pi \in \Pi} v^{\pi}(s)$$