Reinforcement Learning.

Markov's Decision process:

$P(S_{t+1} \mid S_t) = P(S_{t+1} \mid S_1, \ldots S_t)$. - markov

$P_{ss'} = P(S_{t+1} = s' \mid S_t = s)$. - transition prob.

$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} \cdots$

$= \sum_{k=0}^{\infty} \gamma^k R_{k+t+1}$. Returns.

$R_s = E[R_{t+1} \mid S_t]$.

S - states

P - transition prob

R - reward

γ - discount factor.

$\pi(a|s) = P(A_t = a \mid S_t = s)$

policy fn.

State value fn = $V_\pi(s) = E_\pi[G_t \mid S_t = s]$.

(for a given policy).

$V(s) = E[R_{t+1} + \gamma v(S_{t+1}) \mid S_t = s]$.

Bellman.

Scanned with CamScanner

· State- action value fn.

$$q_\pi(s,a) = E_\pi[G_t \mid S_t = s, a_t = a]$$

$$= E_\pi\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \mid S_t = s, a_t = a\right]$$

$$V(s) = R_{s+} \gamma \sum_{s' \in S} P_{ss'} V(s').$$

$$V = R_s + \gamma P V$$

$$(1-\gamma P)V = R_s$$

$$V = (1-\gamma P)^{-1} R_s.$$

Bellman linear eq.

Bellman expectation eq:

$$V(s) = E[R_{t+1} + \gamma V(S_{t+1}) \mid S_t = s].$$

Scanned with CamScanner