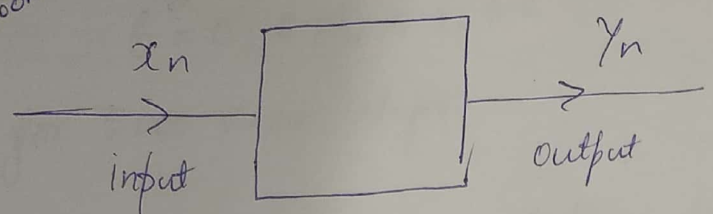
8-Sept-2018
(1 sep notes in ~~Sheets~~ MBRDI-1 Book)



Given (x_n, y_n), $n \geq 0$ tuples & they are revealed one at a time.

Suppose $x_n \in R^n$, $y_n \in R^k$, $m, k > 0$

suppose $f_w : \mathbb{R}^n \longrightarrow \mathbb{R}^k$ where $w = (w_1 \ldots w_d) \in \mathbb{R}^d$

→ f_w is a parametrised class of functions, parameterized by w.

→ f_w could be combinations of sines, coses & polynomials, NN etc

$$f_w(x) = w_0 + w_1 x + w_2 x^2 \ldots + w_d x^2$$

then $w = (w_0, w_1 \ldots w_d)$

<u>Goal:</u> Find the best w — i.e., best explains the output as a function of input.

$$y_n = f_w(x_n) + \varepsilon_n$$ where ε_n is the measurement error.

Now, $\varepsilon_n = y_n - f_w(x_n)$

If we are interested in min Mean Sq error, then

$$\frac{1}{2} E \|\varepsilon_n\|^2 = \frac{1}{2} E \left[\| y_n - f_w(x_n) \|^2 \right]$$

Objective f^n:

$$J(w) = \frac{1}{2} E\left[\| y_n - f_w(x_n)\|^2\right]$$

where $\|.\|$ is euclidean norm in $\mathbb{R}^k$

If $x = (x_1, x_2, x_3 \ldots x_k)$ then

$$\|x\| = \sqrt{x_1^2 + x_2^2 + x_3^2 \ldots x_k^2}$$

$$\|x^2\| = x_1^2 + x_2^2 + \ldots x_n^2 \text{ ie, } x^T x$$

$Y_n^T f_w = f_w^T Y_n$

Thus $J(w) = \frac{1}{2} E\left[(Y_n - f_w(x_n))^T (y_n - f_w(x_n))\right]$

$$= \frac{1}{2} E\left[Y_n^T Y_n + f_w(x_n)^T f_w(x_n) - 2 f_w(x_n)^T Y_n\right]$$

find w^* that ~~min~~ min $(J(w))$

$\Rightarrow$ find $w^* = \underset{w}{\operatorname{argmin}} J(w)$

$$\nabla_w J(w) = \frac{1}{2} \nabla_w E\left[Y^T Y_n + f_n(x_n)^T f_w(x_n) - 2 f_w(x_n)^T \cancel{Y} Y_n\right]$$

Note: $f : \mathbb{R}^d \to \mathbb{R}$

$\nabla f(x) = (\nabla_1 f(x), \ldots \nabla_d f(x))$

where $\nabla_1 f(x) = \frac{\partial f(x)}{\partial x_i}$

Assume that the gradient & expectation of ~~gradient~~ are inter-changeable. Then,

$$\nabla_w \cancel{E} J(w) = \frac{1}{2} E\left[\nabla_w \left(Y^T Y_{n} + f_n(x_n)^T f_w(x_n) - 2 f_w(x_n)^T Y_n\right)\right]$$

$$= \frac{1}{\cancel{2}} E\left[\cancel{2} \nabla_w f_w(x_n)^T f_w(x_n) - 2 \nabla_w f_w(x_n)^T Y_n\right]$$

$$= \cancel{E \nabla_w} E\left[\nabla_w f_w(x_n)\left(f_w(x_n) - Y_n\right)\right]$$

$$= -E\left[\nabla_w f_w(x_n)^T (Y_n - f_w(x_n))\right]$$

Suppose we know the system dynamics, so the expectation is computable. Then the usual gradient descent would give

$$W_{n+1} = W_n - \nabla_{W_n} J(W_n)$$
$$= W_n + E\left[\nabla_W f_W(X_n)^T (Y_n - f_{W_n}(X_n)) \right]$$

Suppose we do not know the system dynamics $\therefore$ we dont know $E[\cdot]$, but we have acess to samples

$$\nabla_{W_n} f_{W_n}(X_n)^T (Y_n - f_{W_n}(X_n))$$

Then we have,

$$W_{n+1} = \cancel{W_n + \nabla W_n f_{W_n}(X_n)^T (Y_n - f_{W_n}(X_n))} \quad \text{--- } \cancel{1}$$
$$W_n + \nabla_{W_n} f_{W_n}(X_n)^T (Y_n - f_{W_n}(X_n)) \quad \text{--- } 1$$
$$W_{n+1} = W_n \quad \text{--- } 2$$

1 would not converge, since it iid (X_n) & represents an random walk.

Suppose $1 > a(n) > 0$ closer to zero

multiply 1 with $a(n)$ & 2 with $1 - a(n)$ and add

$$a(n)(W_{n+1}) + (1 - a(n)) W_{n+1} = a(n)\left(W_n + \nabla_{W_n} f_{W_n}(X_n)^T (Y_n - f_{W_n}(X_n)) \right)$$
$$+ (1 - a(n)) W_n$$

~~Dont Write~~

$$\Rightarrow w_{n+1} = w_n + a(n)\, \nabla_{w_n} f_{w_n}(x_n)^T \left(y_n - f_{w_n}(x_n)\right), \; n \geq 0$$

$$= w_n + a(n)\, E\left[\nabla_n f_{w_n}(x_n)^T \left(y_n - f_{w_n}(x_n)\right)\right]$$

$$+ a(n)\Big(\nabla_w f_{w_n}(x_n)^T \left(y_n - f_{w_n}(x_n)\right)$$

$$- E\left[\nabla_w f_{w_n}(x_n)^T \left(y_n - f_{w_n}(x_n)\right)\right]\Big)$$

Now, $w_{n+1} = w_n - a(n)\, \nabla_{w_n} J(w_n) - a(n)\, M_{n+1}$, —— $(*)$

Where $M_{n+1} = -\nabla_{w_n} f_{w_n}(x_n)^T \left(y_n - f_{w_n}(x_n)\right)$

$+ E\left[\nabla_{w_n} f_{w_n}(x_n)^T \left(y_n - f_{w_n}(x_n)\right)\right]$ $\quad M_{n+1}$ - is mean zero

We replace $\sum_n a(n) = \infty$, $\sum_n a(n)^2 < \infty$

We start (t) in w_0, i.e.,

$$w(0) = w_0 + \int_0^t \nabla_w J(w(z))\, dz$$

As $t \to \infty$,

$$w(t) = H \equiv \{ w \mid \nabla_w J(w) = 0 \}$$

We can then argue that the algorithm will asymptotically converge to H with prob $\rightarrow 1$

Stochastic Approx Scheme:

$$x_{n+1} = x_n + a(n)\left(h(x_n) + M_{n+1}\right), \; n \geq 0$$

~~NAME~~ ~~DATE~~ ~~TIME~~ ~~PLACE~~

~~ADITYA E RAJU~~

~~VIKRAM NATH~~

~~K R VARUN~~

Consider a scenario where noise enters the objective as an arguement $g(x_n, \eta_n)$ where η_n $n \geq 0$ are i.i.d.

$$h(x_n) = E_\eta \, g(x_n, \eta_n)$$

$$M_{n+1} = g(x_n, \eta_n) - E[g.(x_n, \eta_n)]$$

This is similar ~~to~~ ~~for~~ noise can be ~~considered as the~~ represented as previous case. thus $g(x_n, \eta_n)$ is a special case.

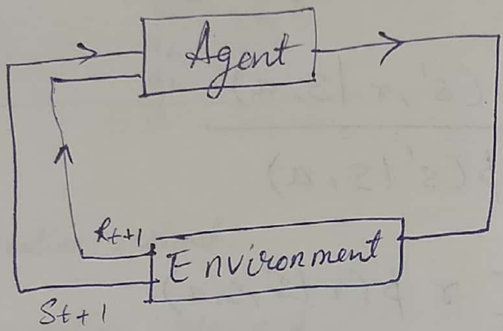
Scanned by CamScanner

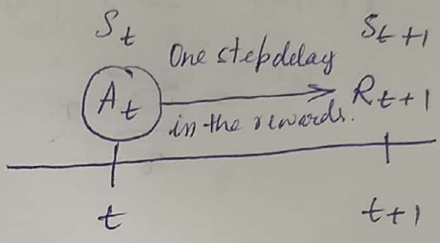
CHAPTER 3:

Finite Markov Decision Process

Decision Maker - Agent

Agent interacts with Environment





R_{t+1}, S_{t+1} depend on S_t, A_t together

The trajectory evolves as follows.

$S_0, A_0, R_1, S_1, A_1, R_2, S_2, A_2, R_3 \ldots$

We assume that sets of states & actions & rewards are all finite.

Let $p(s', r | s, a) \triangleq P(S_{t+1} = s', R_{t+1} = r | S_t = s, A_t = a)$ $\forall s, s' \in S$, $r \in R$, $a \in A(s)$

Set $S \equiv$ set of all states. In state s_0, the set of feasible actions $\equiv A(s_0)$

Now, $\sum_{s' \in S} \sum_{r \in R} p(s', r | s, a) = 1$

joint prob

$$p(s'|s,a) \triangleq P(S_{t+1}=s' | S_t=s, A_t=a)$$

$$= \sum_{r} P(S_{t+1}=s', R_{t+1}=r | S_t=s, A_t=a)$$

$$= \sum_{r} p(s', r|s,a)$$

Let $r(s,a) \triangleq E[R_{t+1} | S_t = s, A_t = a]$

$$= \sum_{r} r \frac{p(s', r|s,a)}{p(s'|s,a)}$$

then $r(s,a) = \sum_{r} r \, p(r|s', s,a)$

Example Movement of robotic arm to pick & place objects.

Learning agent controls motor directly & has all recent information required. (Current positions / velocity / mechanical linkage)

Actions: Voltages applied to each motor at each joint

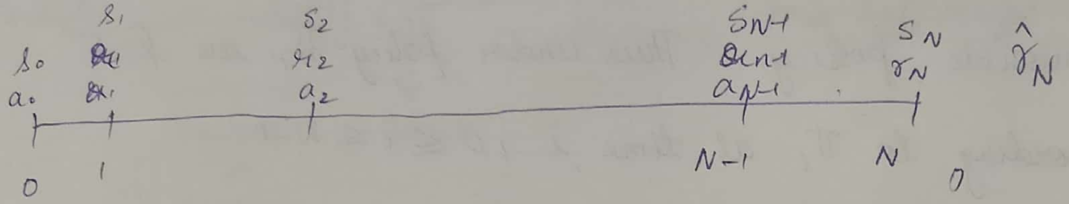
States: Readings of joint angles & velocities.

Rewards:
$$\begin{cases} +1 & \text{if objects is correctly placed.} \\ 0 & \text{otherwise} \\ \text{small -'ve reward if motion is jerky.} \end{cases}$$

Exercise:

Device three ~~examp~~ example tasks that fit into the MDP framework. Identify in each case the state, actions & rewards.

Finite Horizon MDPs



N: termination instant.

r_N: termination reward.

transition probabilites of states

$$p(s' \mid s, a)$$

$$\sum_{s'} p(s' \mid s, a) = 1$$

Expected single state reward

$$\underbrace{r_i(s,a)} \triangleq E\left[R_{i+1} \mid S_i = s, A_i = a\right] \longrightarrow \text{for non terminal states.}$$

where i denotes i^{th} period.

Let $\hat{r}_N(s) \triangleq E\left[\hat{R}_N \mid S_N = s\right] \longrightarrow$ for terminal state.

$r_N(s)$ — expected terminal reward in state s

$r_i(s,a)$ — expected single state reward s.t. - Such that

Suppose that,

Let $\Pi = \{ \pi_0, \pi_1 \dots \pi_{N-1} \}$

Where $\forall i=0, \dots N-1$

$\pi_i : S \longrightarrow A$ Such that $\pi_i(S) \in A(S)$

such that $s \in S$, $\pi_i(s) \in A(s)$

Π: Admissible policy. Thus under policy Π, we pick actions according to π_i at time i, $0 \le i \le N-1$

Suppose ~~$J_\pi(s_0) = E [\sum_i x$~~

$$J_\pi(s_0) = E \left[\sum_{i=0}^{N} r_i(s_i, \pi_i(s_i)) + \hat{r}_N(S_N) \right)$$

Goal: Maximize $J_N(s_0)$ wrt π

Find π^* s.t.

$$J^*(s_0) \triangleq \max_\pi J_\pi(s_0) = J_{\pi^*}(s_0)$$

Result: $J^*(s_0)$ can be obtained by following algorithm.

$$V_N(S_N) = \hat{r}_N(S_N)$$

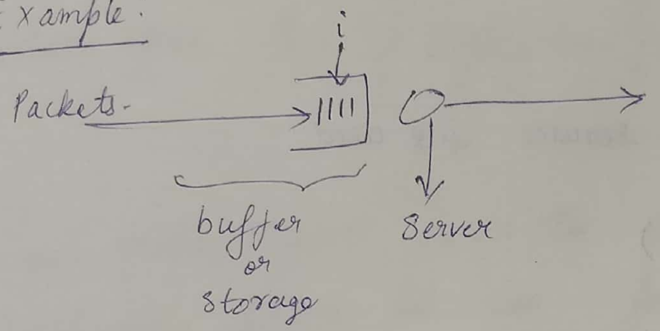
$$V_k(S_k) = \max_{a_k} \sum_k p(s_k, a_k, s_{k+1}) \left(r_k(s_k, a_k) + V_{k+1}(S_{k+1}) \right) + \hat{r}_N(S_N)$$

$$\underbrace{k = N, N-1 \dots 1, 0}_{\text{going backwards in time}}$$

$$\boxed{V_0(S_0) = J^*(S_0)}$$

in the above result, S_0 is not zero state, it is applicable to all the states.

Example:



m= buffersize

Server can provide fast service or slow service. Fast servic comes at a cost C_f. Slow service comes at a cost C_s such that $C_f > C_s$

Let P_m = prob of m arrivals in a ~~block~~ slot.

- A packet can take 1 or more slots for service and ends service just before end of a slot.

- If fast service is used, prob of service completion by the end of slot $= P_f$

- If slow service is used, prob of service completion by the end of slot $- P_s$

$$P_s > P_f$$

Darwin | Sunny + C | Sunny + C | Sunny + C

Suppose $r(i)$ = cost of holding "i" packets in a system
Where i is no of packets in the system.

- Single stage cost when fast server is used
  $g(i, fast) = C_f + r(i)$

- Single stage cost when slow service is used
  $$g(i, slow) = C_s + r(i)$$

$$p(m \mid 0, fast) = \sum_{n=m}^{\infty} p_n$$ — 0 the given that zero packet intially

~~$p(m|1)$~~ $p(m \mid 1, fast) =$ Case 1: Customer in service leaves system by end of slot.

p_f

Case 2: Customer does not leave

$$(1-p_f)\sum_{i=m-1}^{\infty} p_i$$

$(m-1)$ arrivals $(1-p_f)$