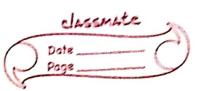
Mudit Gang 2016057



	Chieffy or = and of free
dus I	Initialize:
Alle	M(s) EA(s) (arbitrarily), 45ES
	Q(s.a) & IR (arbitrailly) . YSES YaEA
	Returns sum (5,a) ER (initialize all values to 0)
	Returns-count(s,a) ER (initialize all values to0)
	loop forever (for each episode)?
	choose so Es, A EA (so) randomly such that
	all pairs have probability >0
	generate an episode from So, Ao following
	TI: So As River ST-1 RT-1 RT
	$G \leftarrow O$
	loop for each step of the episode t= T-1.0
	Loop for each step of the episode, t= T-1.0
	Returns_sum(S_t , A_t) $t = G_1$
	Retrons _ count (St, At) + = 1
	O(Sto A) = Returns sum (St, A)/Returns
	$O(S_t, A_t) = Returns sum(S_t, A_t)/Returns$ $II(S_t) = argmax (O(S_t, a)) count(S_t, A_t)$
	$\Pi(S_t) = \alpha i g \alpha (S(t))$
	T: HA
7	AM - HAMP

Ans 2	$q_{\Pi}(s,a) = 1 + V_{\Pi}(s'a)$
	= 97(s,a) + a (Git- 97(s,a))
	TEST ACT (WESTAMILY)
	So Qualitation A 3 acres
13 43	halou ha a find on the share
(10 H 60)	Remon ount (s, a) to R (in) size all val
	fire forescer (for cach eps on) 3
Host nou	14005 S. 65, A. E. A. C. O. Standard
	all pairs have presidenties
(grown)	
	2 anii - 2 9 . 10 2 2 77
	7n 036
	location was the second
	A HAY - JA
	Reference survise. A. D. T. = (mg
Ams 3	\$ 9π(s,a) = Eπ(Gt St=s, At=a)
35447	Alla Blowns samuels = C.A. 21A
A	Now Property Some And Sold Sies Area
	Now, Pr (Men Stri, Att Sylst = S, At= a, Att: T-1~1)
	The San Action
	= P(Stri S,a) Ti(Atti STri) Ti(AT-1 STri)
	* P(8T) AT-1, ST-1)
	T-I)

=) Pr = TT TT (AKISK) P(SK+1) SK, AK)
K=++1 => Pt+1: T-1 = TT TI(AKISK Also, t+1: T-1 can be written as t ∈ V(s,a) where V(s,a) is trajectory after starting from 5, a & St: T(t)-1 Git t∈ Y(s,a) J. St:T(t)-1 ter(s,a)