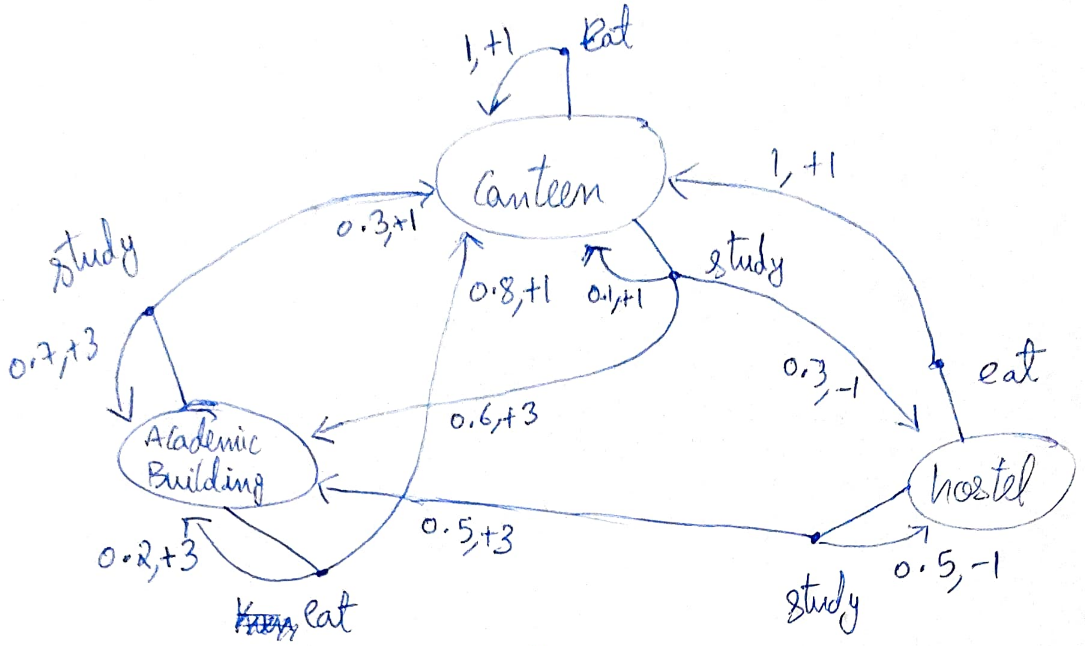
28/8/24

Assignment - I

SUCHETAN G.U.
21342

Q1)



MDP diagram

s	a	s'	$P(s',r \mid s,a)$	$\gamma(s,a,s')$
Canteen	study	Canteen	0.1	+1
Canteen	study	hostel	0.3	-1
Canteen	study	Academic Build.	0.6	+3
Canteen	eat	Canteen	1	+1
hostel	study	hostel	0.5	-1
hostel	study	Academic Build.	0.5	+3
hostel	eat	canteen	1	+1
Academic Build	study	Academic Build	0.7	+3
Academic build	study	canteen	0.3	+1
Academic Build	eat	Academic Build	0.2	+3
Academic Build	eat	canteen	0.8	+1

other transitions have Probability = 0 and hence no reward.

⇒ Both Value and policy iteration show Same trend in state -Value function,

⇒ V (academic building) > V (canteen) > V (hostel) as expected because of the rewards associated with these places show same pattern.

⇒ The optimal policy in both cases was to 'study' in any locations when possible.