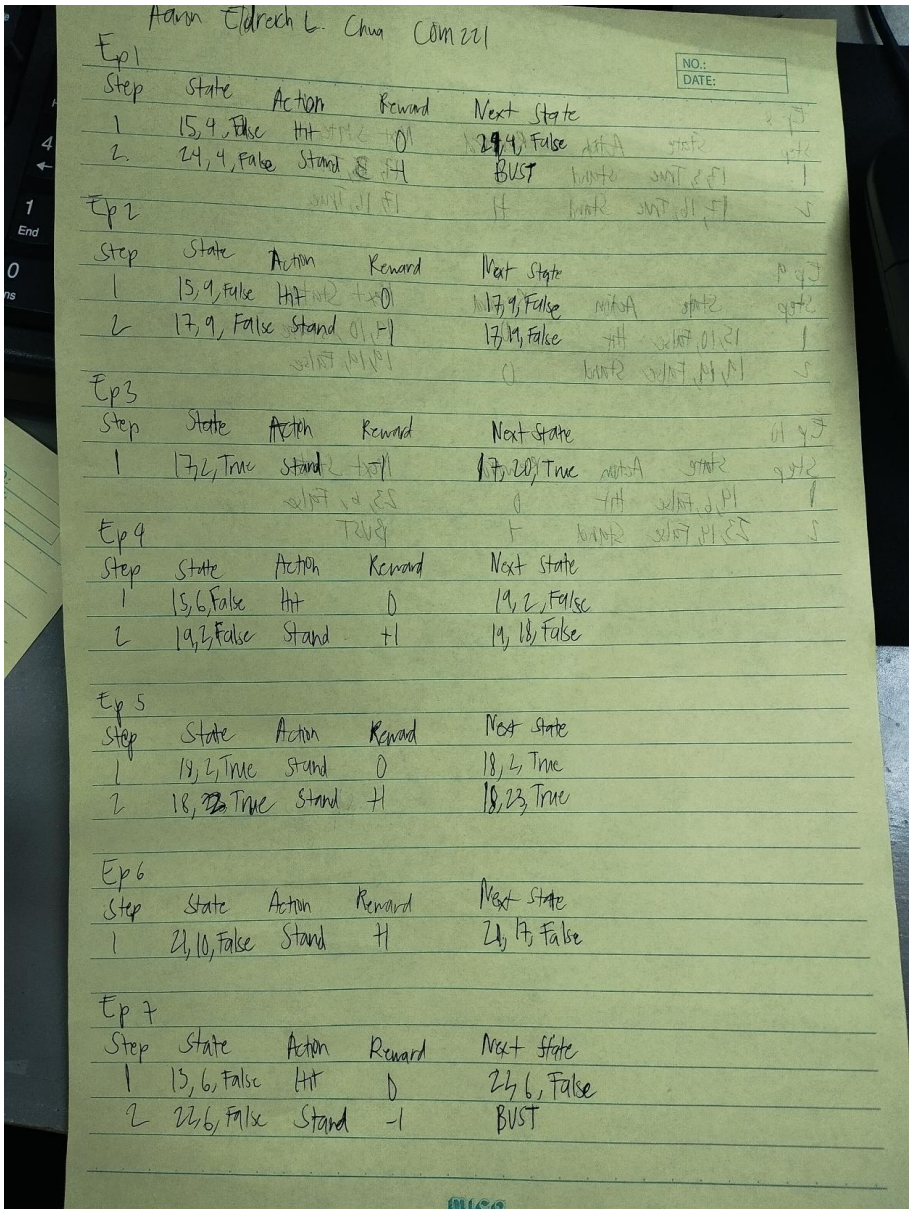
Chua, Aaron Eldreich L. COM221 Exercise 3



Aaron Eldreich L. Chua COM221

NO.:
DATE:

Ep 1

Step	State	Action	Reward	Next State
1	15, 9, False	Hit	0	24, 9, False
2	24, 9, False	Stand	-1	BUST

Ep 2

Step	State	Action	Reward	Next State
1	15, 9, False	Hit	0	17, 9, False
2	17, 9, False	Stand	-1	17, 19, False

Ep 3

Step	State	Action	Reward	Next State
1	17, 2, True	Stand	+1	17, 20, True

Ep 4

Step	State	Action	Reward	Next State
1	15, 6, False	Hit	0	19, 2, False
2	19, 2, False	Stand	+1	19, 18, False

Ep 5

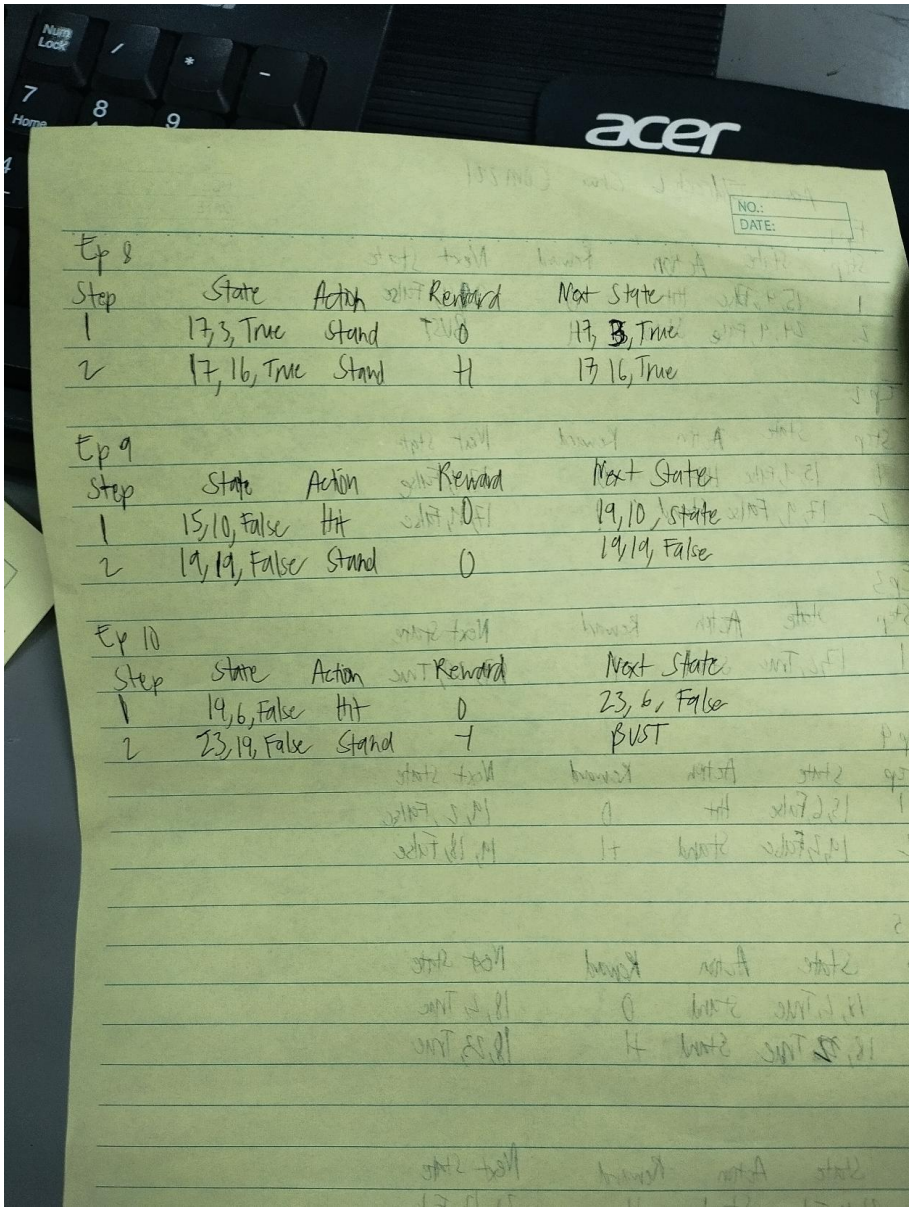
Step	State	Action	Reward	Next State
1	18, 2, True	Stand	0	18, 2, True
2	18, 22 True	Stand	+1	18, 23, True

Ep 6

Step	State	Action	Reward	Next State
1	21, 10, False	Stand	+1	21, 17, False

Ep 7

Step	State	Action	Reward	Next state
1	15, 6, False	Hit	0	21, 6, False
2	21, 6, False	Stand	-1	BUST



Ep 8

Step	State	Action	Reward	Next State
1	17, 3, True	Stand	0	17, 3, True
2	17, 16, True	Stand	+1	17, 16, True

Ep 9

Step	State	Action	Reward	Next State
1	15, 10, False	Hit	0	19, 10, False
2	19, 19, False	Stand	0	19, 19, False

Ep 10

Step	State	Action	Reward	Next State
1	19, 6, False	Hit	0	23, 6, False
2	23, 19, False	Stand	-1	BUST

Exercise 5

Aaron Eldrich L. Chua COM221

NO.:
DATE:

Part B: Monte Carlo Update

Eps.

18,2,True

Part B: Monte Carlo Update

Ep 5:

State	Return G	Visit Count	Old V(s)	New V(s)
18,2, True	+1	1	0	1
18,23, True	+1	1	0	1

State 1: New V(s)

$0 + \frac{1}{1}(1-0) = 1$ //

State 2: New V(s)

$0 + \frac{1}{1}(1-0) = 1$ //

Part C: Temporal Difference Update

a. State = (18, 2, True)

b. Reward = 0

c. Next State = (18, 23, True)

d. Update: V(18, 2, True) = $0 + 0.5(0 + 0 - 0) = 0$ //

Step	State s	Reward r	Next State s	Old V(s)	New V(s)
1	18, 2, True	0	18, 23, True	0	0
2	—	—	—	—	—

Update: V(18, 23, True) = $0 + 0.5(1 + 0 - 0) = 0.5$ //

Step 2:

Step	State	Reward	Next State	Old	New
1	(18, 2, True)	0	18, 23, True	0	0
2	(18, 23, True)	+1	WIN	0	0.5

Step 3:

Update: $V(18,2,\text{True}) = 0 + 0.5(1 + 0.5 - 0) = 0.25$

Step	State	Reward	Next state	Old	New
1	18, 2, True	0	18, 23, True	0	0.25
2	18, 23, True	+1	WIN	0	+0.5