Q(s,a) = Q(s,a) + OC[r+&maxQ(s',a') - Q(s,a)]

-		al
51	0	0
SL	0	0
53	10	10

5:51

Q(\$1, A1)=0+0.5[0.5+0.5(0)-0]=0.25

a; al

r: 0.5

5'. 52

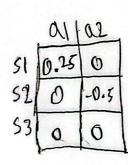
1	140	ar
51	0.3	0
52	0	0
53	0	0

Q(52,02) = 0+0.5[-1+0.5(0)-0] = -0.5

0:02

1: -1

51: 533



Offinal Policy

- of 0.5 or 10 from 52 +51 by al
- from 52, just action is alt 51 because it has a remark of 10 to from 51+52 to through a 10011. So you get remarks cool of 10 and 0.5 inawley.
- has a loop from 52 + 51 resulting in a to set loop which years of to. It also constant remark of 0.5 and 10