...../...../.....

RL

Task 1:

a) S: $\{5, 4, 3, 2, 1\} \cup \{Parked, Failed\}$

A: P (try to park in current parking space)
D (Drive on to next space)

p: if $s' = Parked \rightarrow p(s', 1 \mid s, T) = p$

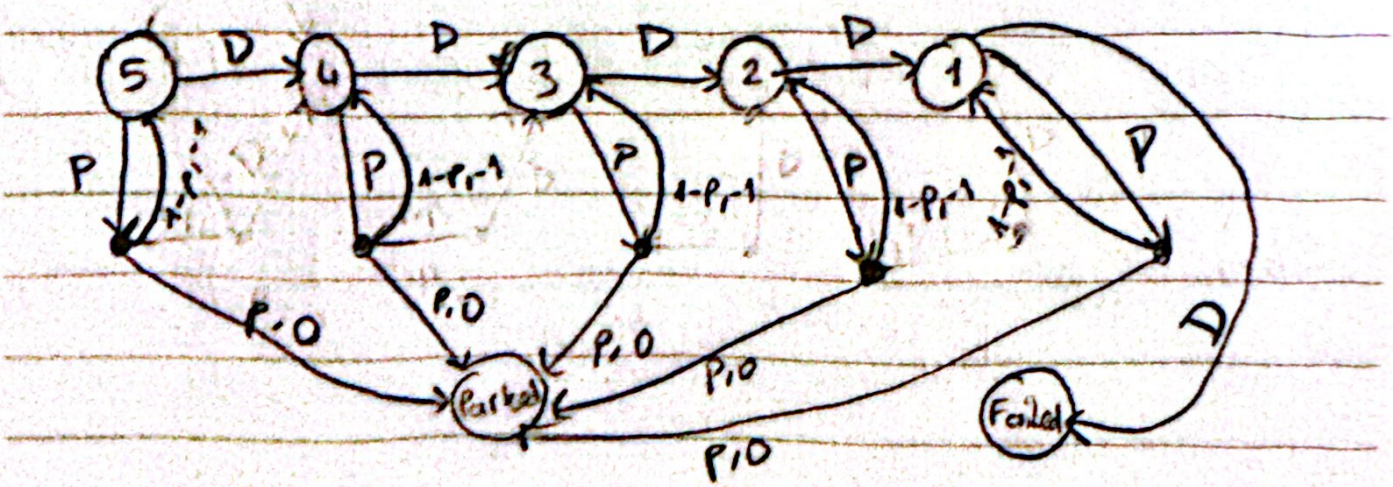
if $s' = s-1 \rightarrow p(s', -1 \mid s, D) = 1$

$\rightarrow p(s', -1 \mid s, P) = 1-p$

if $s = 1$ and $a = D \rightarrow p(Failed, -10 \mid 1, D) = 1$

R: $\{+1, -1, -10\}$

b)



c) Discount can be applied because we want to park as quickly as possible and therefore future rewards should be decreased by a discount factor.

Task 2) Since the next state is determined not only from the current state but also the previous state (last 2 states), Markov property is not fullfilled.

Task 3)

Disprove

Consider $S = \{s_1, s_2\}$

$A = \{a_1, a_2\}$

In s_1

→ a_1 : transitions to s_2 with $r = 1$

a_2 : " " " "

In s_2

→ both actions go back to s_1 with a reward of 0

So $v^*(s_1) = 1$, $v^*(s_2) = 0$

and these don't change with different policies

$\pi^*(s_1) = a_1 = a_2$

$\pi^*(s_2) = a_1 = a_2$

optimal deterministic policy is not unique.