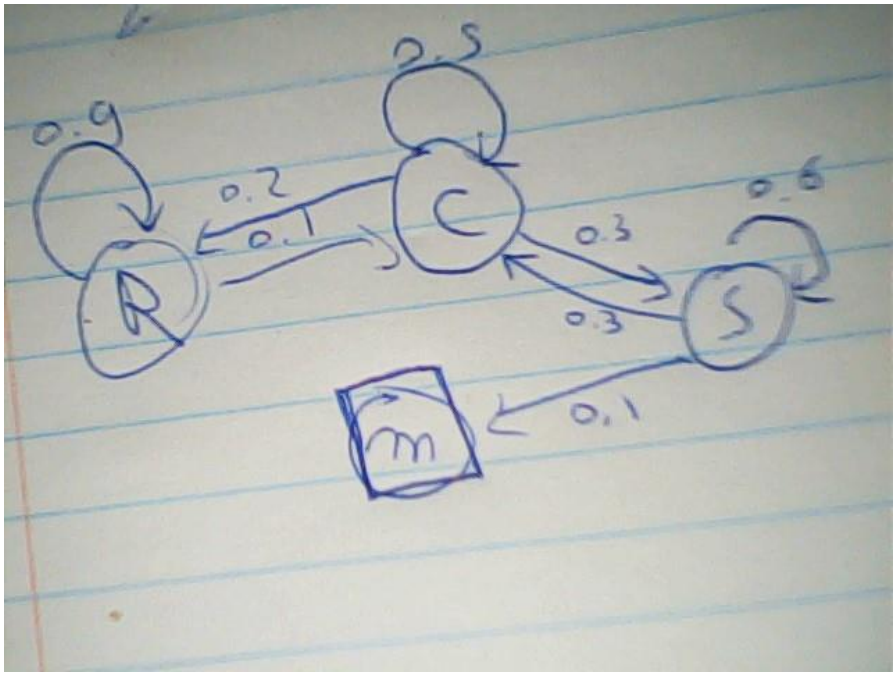
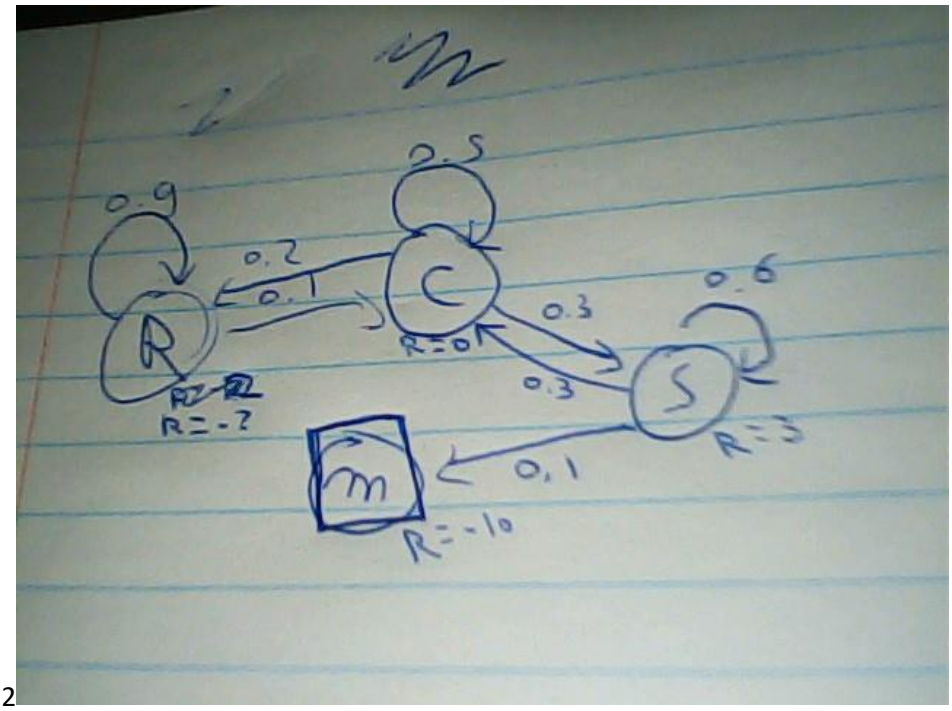
Inleveropgave 1: Model-based Prediction and Control

1.1



1.2



1.3

Sample 1:

Rain > Rain > Rain > Cloudy > Sunny > Sunny > Meteor

Gt = -2 + (-2*1) + (-2*1) + (0*1) + (3*1) + (3*1) + (-10*1) = -10

Sample 2:

Sunny > Cloudy > Sunny > Cloudy > Rain > Rain > Cloudy > Sunny

Gt = 3 + (0*1) + (3*1) + (0*1) + (-2*1) + (-2*1) + (0*1) + (3*1) = 5

1.4

$$
\begin{aligned}
v_\pi(s) &\doteq \mathbb{E}_\pi[G_t \mid S_t = s] \\
&= \mathbb{E}_\pi[R_{t+1} + \gamma G_{t+1} \mid S_t = s] && \text{(from (3.9))} \\
&= \mathbb{E}_\pi[R_{t+1} + \gamma v_\pi(S_{t+1}) \mid S_t = s] && (4.3)
\end{aligned}
$$

Rain	0	-1.8	-3.37
Cloudy	0	0.5	0.32
Sunny	0	0.8	1.37
Meteor	0	0	0

1.5

1) Met een gamma die bijna 1 is of 1 is, krijg je “far-sighted” evaluations, dan kan een probleem zijn als je direct een reward wilt

2) kan zorgen voor een infinite loop

2

iter	s1	s2	s3
0	0	0	0
1	-0,1	-0,1	0
2	-0,2	-0,2	0
3	-0,3	-0,3	0
4	-0,4	-0,4	0
5	-0,5	-0,5	0
6	-0,6	-0,6	0
7	-0,7	-0,7	0
8	-0,8	-0,8	0
9	-0,9	-0,9	0
10	-1	-1	0
11	-1,1	-1	0
12	-1,1	-1	0
13	-1,1	-1	0
14	-1,1	-1	0
15	-1,1	-1	0
16	-1,1	-1	0

Hierna heeft het geen nut meer om te runnen, de values veranderen niet meer.