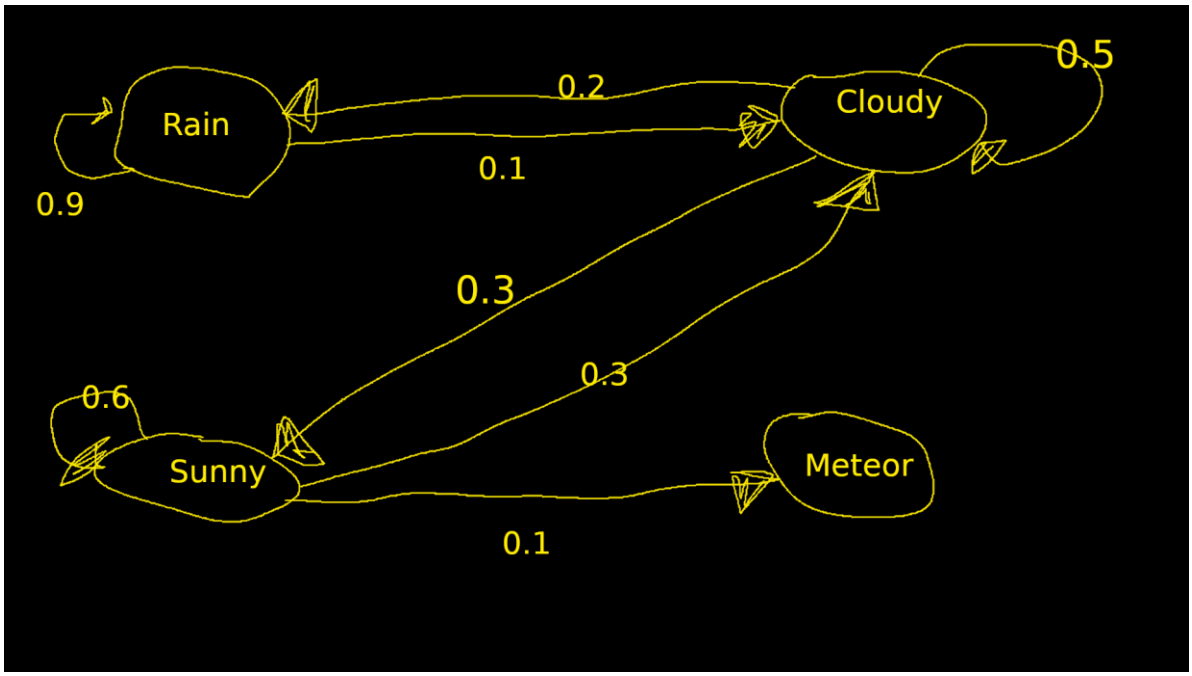
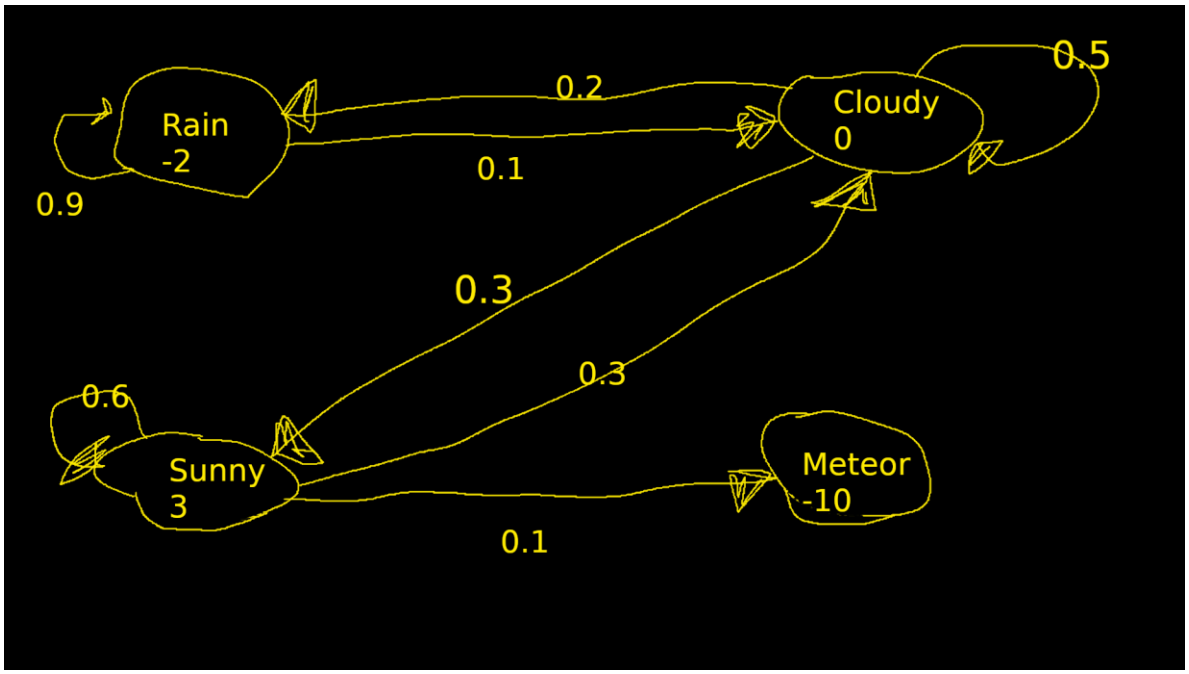
Inleveropgave 1: Model-based Prediction and Control

1.1



1.2



1.3

rain (start) -> cloudy -> cloudy -> sunny -> meteor

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

= 0 + 0 + 3 - 10

= -7

rain (start) -> rain -> cloudy -> sunny -> sunny-> meteor

Gt = (1^0 * -1) + (1^1 * 0) + (1^2 * 3) + (1^3 * 3) + (1^4 * -10)

= -1 + 0 + 3 + 3 - 10

= -5

1.4

Rain	Cloudy	Sunny	Meteor
0	0	0	0
-1.8	0.5	0.8	0
-3.37	0.63	1.43	0

1.5

Een discount factor van 1 kan meerdere problemen met zich meebrengen. Twee hiervan zijn:

1. Er wordt minder gekeken naar directe rewards

2. Er wordt veel weight gegeven aan latere rewards

2

s0	s1	s2 (eind)
0	0	0
-0.1	-0.55	0
-0.65	-1	0
-1.1	-1	0
-1.1	-1	0

Gestopt omdat de values zich herhalen, de optimale values zijn bereikt.