Machine data and learning

Assignment 5 part 1

Yash Bhansali
2018101068

P(Observation = Red | State = Red)	0.85
P(Observation = Green | State = Green)	0.9

Initially, the agent knows that it is in one of the red states i.e S1, S2 or S5.

Initial belief state = `[1/3,1/3,0,0,1/3]`

P[success] = 0.71

B[s] = (O(s', a, o)ΣT(s, a, s') b(s))Pr (o | a, b)

O(s', a, o) represents probability of observing **o** after action '**a**' was taken to reach **s'** i.e. **Pr (o | s, a)**

T(s, a, s') represents probability of going to state **s** after action '**a**' was taken on **s'** i.e. **Pr(s | s', a)**

For calculating new belief state
b[s] = Pr (o | s, a))ΣPr(s | s', a) * b(s')

The above is an unnormalised probability, we will normalise it by dividing each probability by the sum of all the probabilities.

NOTE: And in our case, Pr (o | s, a) is not dependent on action.

Agent took the action Right and observed Red .

b[s1]=0.85[0.29 * 0.333 + 0.29 * 0.333 + 0 * 0 + 0 * 0 + 0 * 0.333]=0.164169

b[s2]=0.85[0.71 * 0.333 + 0 * 0.333 + 0.29 * 0 + 0 * 0 + 0 * 0.333]=0.2009655

b[s3]=0.1[0 * 0.333 + 0.71 * 0.333 + 0 * 0 + 0.29 * 0 + 0 * 0.333]=0.023643

b[s4]=0.1[0 * 0.333 + 0 * 0.333 + 0.71 * 0 + 0 * 0 + 0.29 * 0.333]=0.009657

b[s5]=0.85[0 * 0.333 + 0 * 0.333 + 0 * 0 + 0.71 * 0 + 0.71 * 0.333]=0.2009655

sum of unnormalised belief state = 0.5994

New belief state = **[0.27388, 0.33527, 0.03944, 0.01611, 0.33527]**

Agent took the action Left and observed Green .

b[s1]=0.15[0.71 * 0.27388+ 0.71 * 0.33527 + 0 * 0.03944 + 0 * 0.01611 + 0 * 0.33527]=0.06487

b[s2]=0.15[0.29 * 0.27388 + 0 * 0.33527 + 0.71 * 0.03944 + 0 * 0.01611 + 0 * 0.33527]=0.01611

b[s3]=0.9[0 * 0.27388 + 0.29 * 0.33527 + 0 * 0.03944 + 0.71 * 0.01611 + 0 * 0.33527]=0.09780

b[s4]=0.9[0 * 0.27388 + 0 * 0.33527 + 0.29 * 0.03944 + 0 * 0.01611 + 0.71 * 0.33527]=0.22453

b[s5]=0.15[0 * 0.27388 + 0 * 0.33527 + 0 * 0.03944 + 0.29 * 0.01611 + 0.29 * 0.33527]=0.01528

sum of unnormalised belief state = 0.41861

New belief state = **[0.15497, 0.03849, 0.23363, 0.53637, 0.03651]**

Agent took the action Left and observed Green .

b[s1]=0.15[0.71 * 0.15497 + 0.71 * 0.03849 + 0 * 0.23363 + 0 * 0.53637 + 0 * 0.03651]=0.02060

b[s2]=0.15[0.29 * 0.15497 + 0 * 0.03849 + 0.71 * 0.23363 + 0 * 0.53637 + 0 * 0.03651]=0.03162

b[s3]=0.9[0 * 0.15497 + 0.29 * 0.03849 + 0 * 0.23363 + 0.71 * 0.53637 + 0 * 0.036514]=0.35279

b[s4]=0.9[0 * 0.15497 + 0 * 0.03849 + 0.29 * 0.23363 + 0 * 0.53637 + 0.71 * 0.03651]=0.08431

b[s5]=0.15[0 * 0.15497 + 0 * 0.03849 + 0 * 0.23363 + 0.29 * 0.53637 + 0.29 * 0.03651]=0.02492

sum of unnormalised belief state = 0.51425

New belief state = [0.04006, 0.06149, 0.68603, 0.16394, 0.04846]