Report of Tutorial 5 in BILHR

1. Group Division

Name (ranking wrt first letter of surname)	Division of work
Guo, Qiuhai	Detection of goalkeeper, Speech of NAO, Link between 3 cpps
Hao, Wenhan	Optimization of NAO & ball position, Balance of NAO
Qiu, Tianming	Main body of RL algorithm, Balance of NAO
Xu, Jiachen	Balance of NAO, Kicking& Moving action, Writing of report

Note: All of us four have done this project with equal contribution, especially during the process of debugging. Therefore, it's fairly hard to comment the code with specific coder's name.

2. Plot of cumulative reward

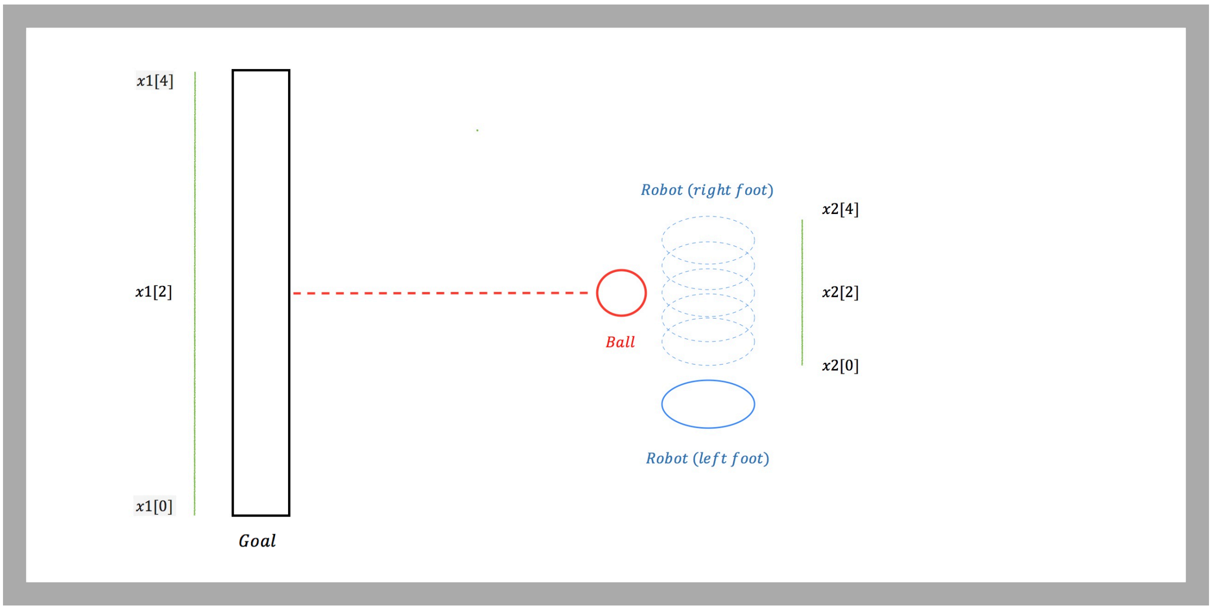

Fig 1. Top view of location during shooting

2.1 Theoretical Analysis:

We arranged the location of robots, gate and ball as described in above figure, the robot keeps balance mainly depending on its left leg and move its right leg to kick, in and out. From their mutual position we can predict that:

1. If the goalkeeper is at position $x1[0]$, the robot won't kick the ball at state position $x2[4]$.
2. If the goalkeeper is at position $x1[1]$, the robot won't kick the ball at state position $x2[3]$.
3. If the goalkeeper is at position $x1[2]$, the robot won't kick the ball at state position $x2[2]$.
4. If the goalkeeper is at position $x1[3]$, the robot won't kick the ball at state position $x2[1]$.
5. If the goalkeeper is at position $x1[4]$, the robot won't kick the ball at state position $x2[0]$.

2.2 Results:

Declaration:

I plot the cumulative reward in different with respect to different actions, as below figure shows, red line represents “kick”, and blue represents “in”, green represents “out”. And **the action with highest reward is the action which will be executed at this state during the test or prediction process.**

Intuitively, the right leg of NAO won’t move in when the right leg of NAO locates at state x2[0] and won’t move out when at state x2[4].

For instance, if the goalkeeper is at x1[0] right now, the robots leg will move in when its leg locates at state x2[4], because from 2.1 we’ve discussed about this theorectical result. On the contrary, it’s possible for robots to kick at any position other than x2[4]. This analysis is as same as the below cumulative reward plot described. The rest figures can also be explained with the same way.

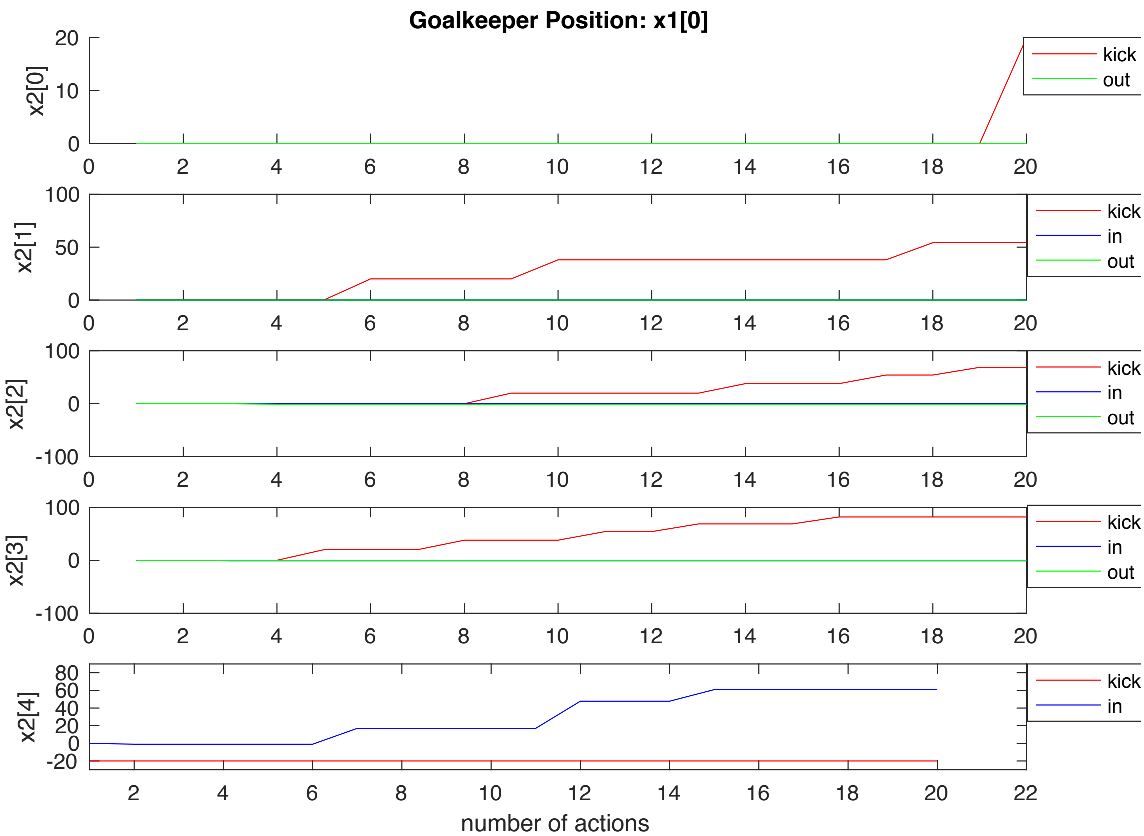

Fig 2. Cumulative rewards when goalkeeper is at state x1[0]

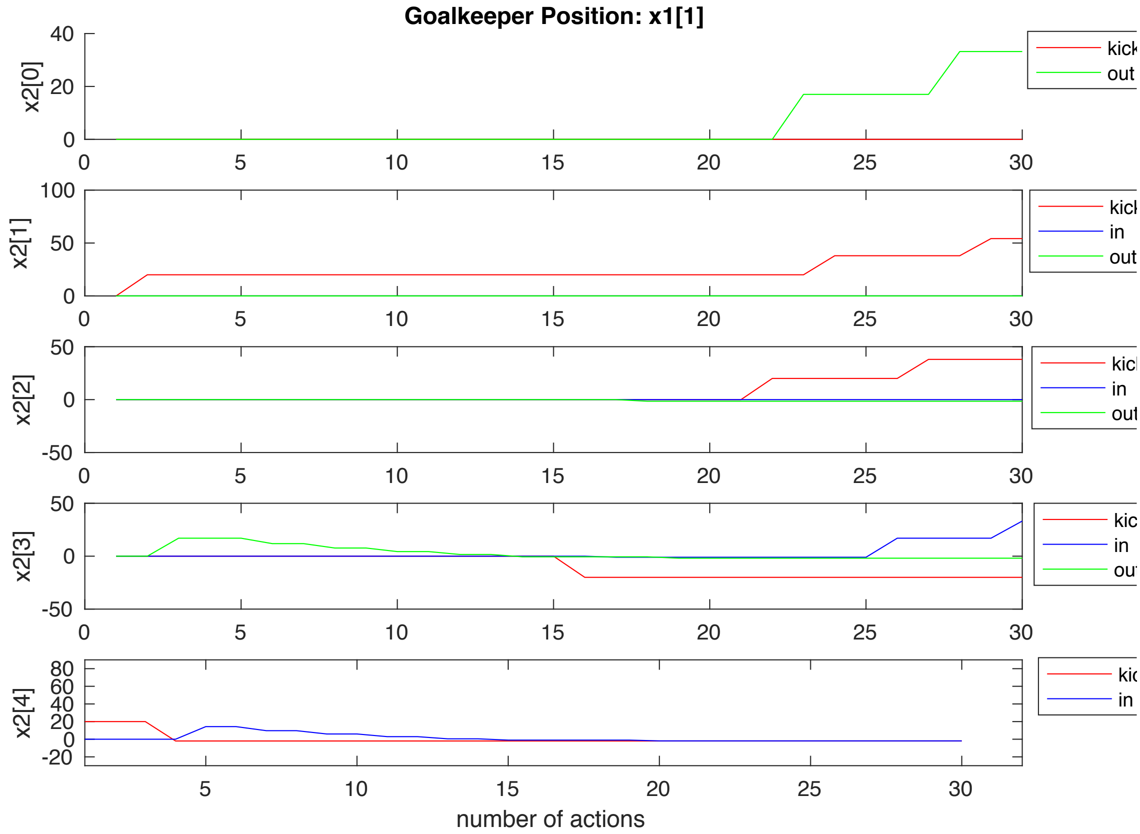

Fig 3. Cumulative rewards when goalkeeper is at state x1[1]

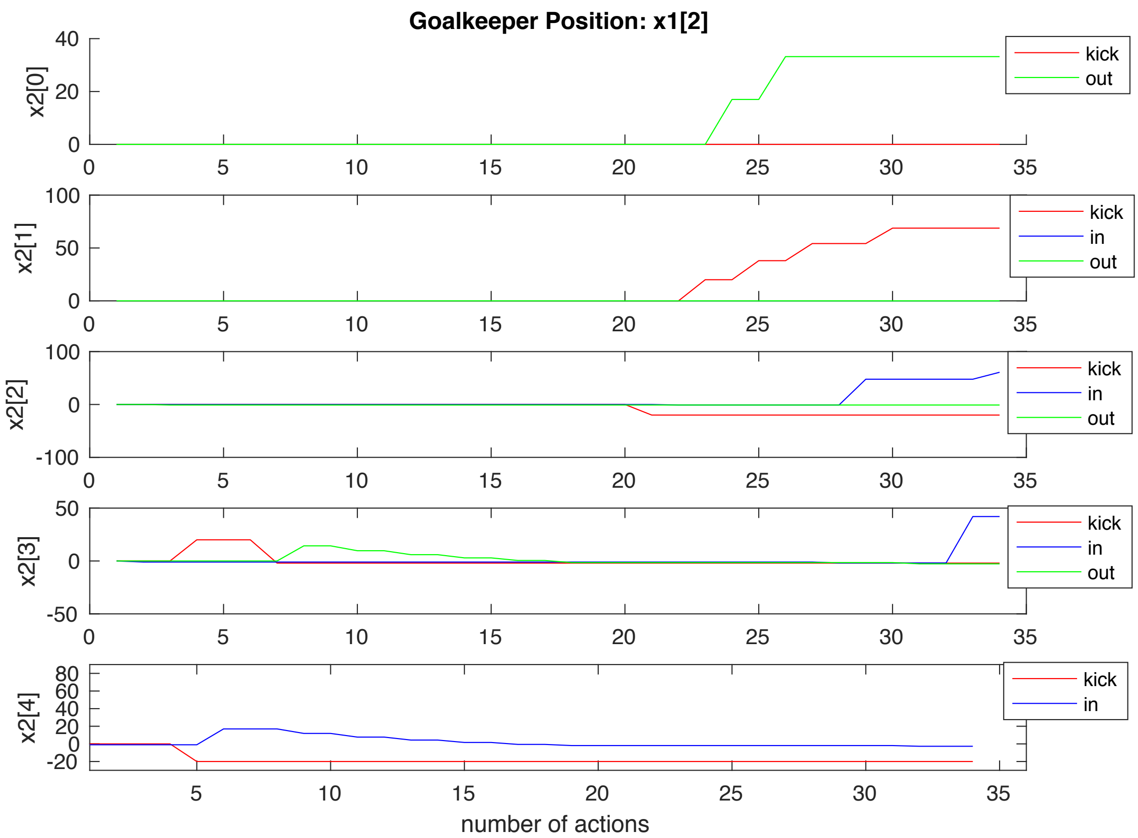

Fig 4. Cumulative rewards when goalkeeper is at state x1[2]

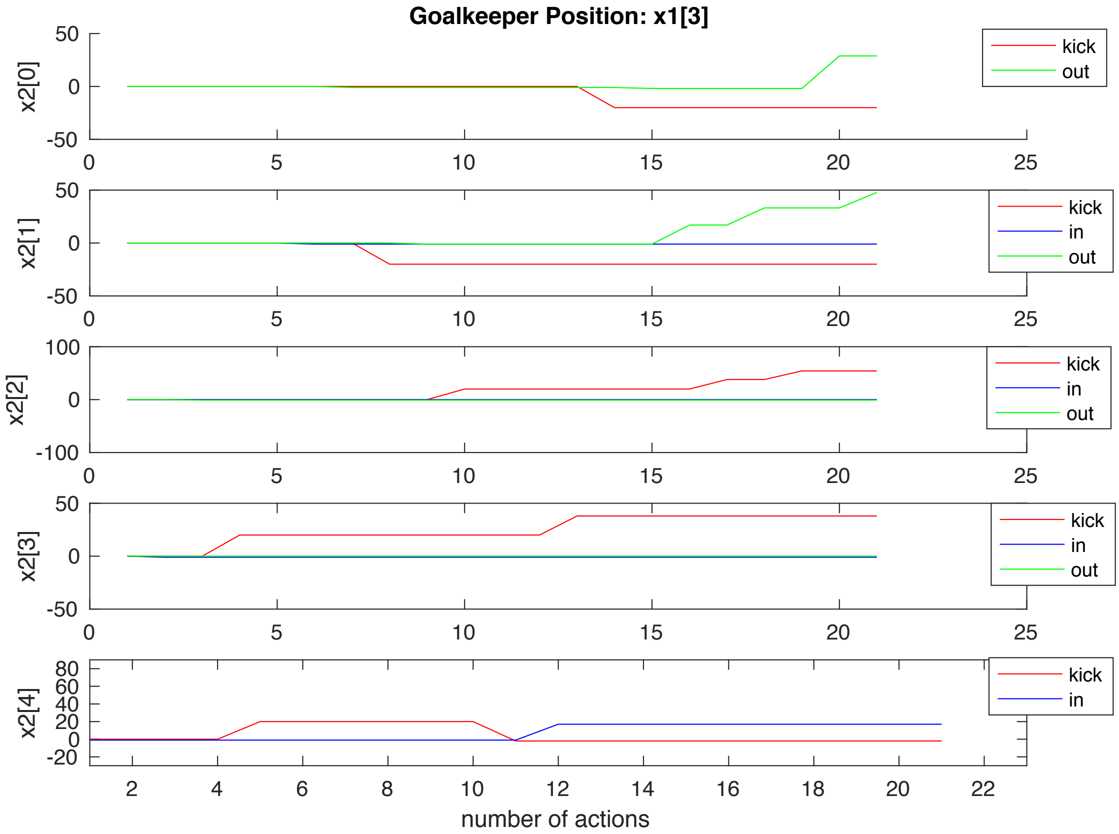

Fig 5. Cumulative rewards when goalkeeper is at state x1[3]

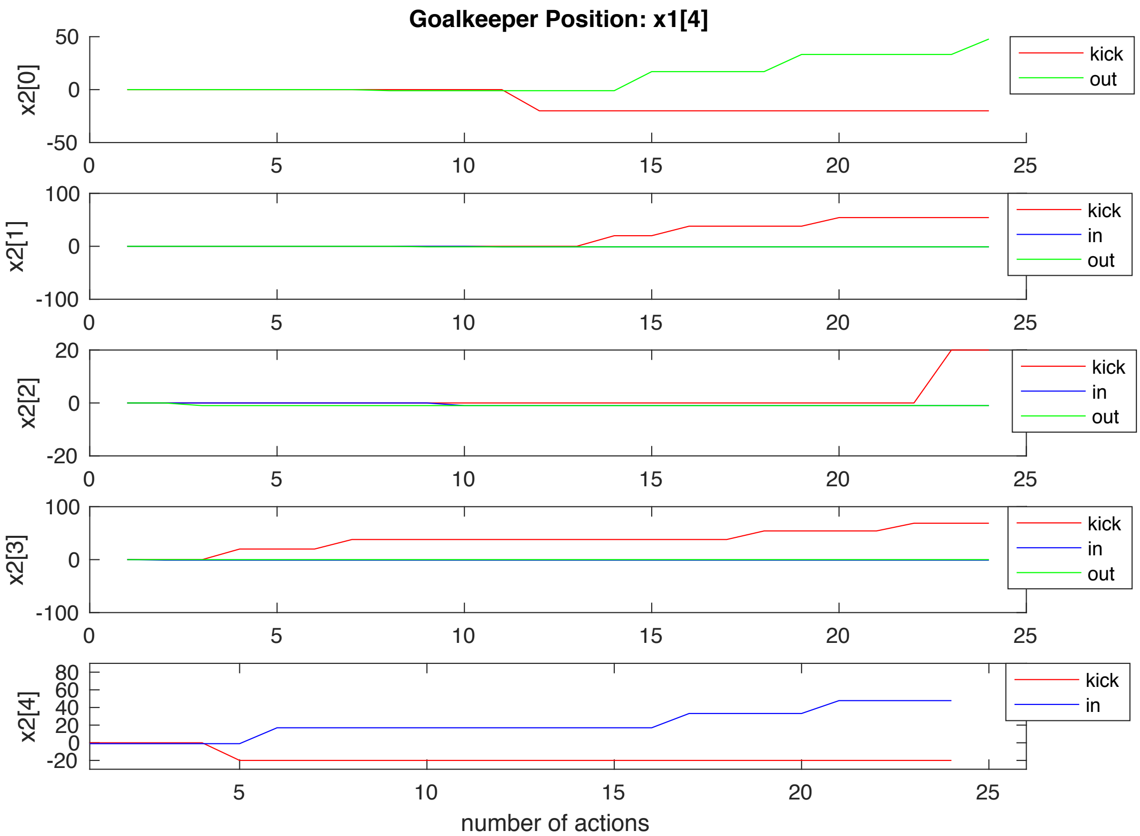

Fig 6. Cumulative rewards when goalkeeper is at state x1[4]

3. Reason for changing penalization (unsuccessful kick) into -20 rather than -5

One of the most significant reasons is by using higher penalization value we can accelerate the convergence rate. Because by increasing the difference between successful & unsuccessful kick, the most stable successful kicking point is faster to be determined. E.g. when we use the -20 negative reward, after 20-30 actions, we can get a converged Q value table. On the opposite, we've tried -5 and it takes much more number of actions to converge.

The second reason is to avoid the unstable successful kick. If the positive and negative value is equal in sense of absoluteness, NAO won't choose to kick if the percentage of successful goal is lower than 50%. However, if the negative reward is -5, even if we have 40% or even 30% successful goal rate, NAO will still choose to kick the ball at this unstable position.

E.g., when the robot kicks the ball at state x2[0] or x2[4], even if we can achieve successful goal, but due to the limit of ground flatness and the highly strict requirement towards the execution accuracy of HipRoll value, these two position could be unstable goal point.

As below figure shows, if we kick the ball at state x2[0] or x2[4], we will use the red circle region and yellow circle region to kick, and slightly change of actual HipRoll value will result in seriously bias of hitting ball point. It is because the boundary of right foot is ellipsoid, which means the curvature is not equal at different point of boundary. Therefore, we need to try our best to avoid these unstable successful kicking.

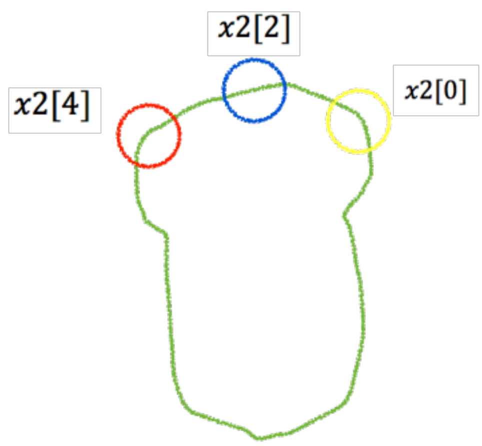

Fig7. The kicking ball region of NAO right foot