Machine Learning

Seminar 9

1. What is the goal of reinforcement learning?

2. What is the difference between supervised learning, unsupervised learning and reinforcement learning?

3. How to describe a reinforcement learning problem? What is environment, state, reward, policy, value?

4. What is optimal policy? Explain the basic idea of the equations below and explain the meaning of all symbols)

$$Q(s,a) = r(s,a) + \gamma \max_{a} Q(s',a)$$

$$Q(s,a) \rightarrow \gamma \mathrm{Q}(s',\mathrm{a}) + \gamma^2 Q(s'',a) \dots \dots \dots \gamma^n Q(s''^{\dots n},a)$$

5. What is Q-table? How Q learning use the table to search?

$$\begin{array}{c} state \backslash action \\ A \\ B \\ C \\ D \\ E \\ F \end{array} \begin{array}{c} \begin{array}{cccccc} A & B & C & D & E & F \end{array} \\ \begin{bmatrix} - & - & - & - & 0 & - \\ - & - & - & 0 & - & 100 \\ - & - & - & 0 & - & - \\ - & 0 & 0 & - & 0 & - \\ 0 & - & - & 0 & - & 100 \\ - & 0 & - & - & 0 & 100 \end{bmatrix} \end{array}$$

6. Lets use the table in Q5 as the instant reward matrix, set gamma=0.8, F as goal state.

- Draw a diagram that represents the action and instant reward.
- Solve the episode with initial state of B.
- Solve the episode with initial state of D.

7. Assume that the final Q table of Q6 is as below. Find the optimal path from C to F.

$$\hat{\mathbf{Q}} = \begin{array}{c} state \backslash action \\ A \\ B \\ C \\ D \\ E \\ F \end{array} \begin{array}{c} \begin{array}{cccccc} A & B & C & D & E & F \end{array} \\ \begin{bmatrix} - & - & - & - & 80 & - \\ - & - & - & 64 & - & 100 \\ - & - & - & 64 & - & - \\ - & 80 & 51 & - & 80 & - \\ 64 & - & - & 64 & - & 100 \\ - & 80 & - & - & 80 & 100 \end{bmatrix} \end{array}$$