Solutions to Reinforcement Learning by Sutton Chapter 9

Yifan Wang

December 2019

Exercise 9.1

Tabular methods are special cases of linear functions when we use one hot encoding for feature function x**, and** $w \doteq [w_1, \ldots, w_N]^\mathsf{T}$ **for total N states.**

For each state s**, we have:**

$$\hat{v}(s, \mathrm{w}) \doteq \mathrm{w}^\mathsf{T}\mathrm{x}(s) \doteq \sum_{i=1}^{d} w_i x_i(s) = w_n * 1 = w_n$$

Gradient vector will coincide with the one-hot encoding:

$$\nabla \hat{v}(s, \mathrm{w}) = \mathrm{x}(s)$$

By applying the above two equations, linear methods can be transformed into equivalent tabular method. For example, consider equation (9.15) in the book:

$$\mathrm{w}_{t+n} \doteq \mathrm{w}_{t+n-1} + \alpha\big[G_{t:t+n} - \hat{v}(S_t, \mathrm{w}_{t+n-1})\big]\nabla \hat{v}(S_t, \mathrm{w}_{t+n-1})\big] \tag{9.15}$$

For any state s, let the position of 1 in $\mathrm{x}(s)$ **be** $p(s)$**, we will have :**

$$\mathrm{w}_{t+n}[p(s)] \doteq \hat{v}_{t+n}(s) \doteq \hat{v}_{t+n-1}(s) + \alpha\big[G_{t:t+n} - \hat{v}(s)\big]$$

By further rewriting equation (9.16) , we thus recover the tabular form of n-step TD method as in Chapter 7.

■

Exercise 9.2

Since set$\{0, 1, \ldots, n\}$contains $n+1$ elements, for each s_i, we have $n+1$ ways of choosing its power. Since we have k states to choose, we will have $(n+1)^k$ choices by law of permutation.

■

Exercise 9.3

Observing the maximum number of choice, we get $n = 2$ and $c_{i,j} \in \{0, 1, 2\}$

■

Exercise 9.4

Any tiles that gives denser representation to the important dimensions will do. For example, rectangular tiles that are shorter in its important dimension or denser strip tiles across it. ■

Exercise 9.5

It is an open question but to make an appropriate learning for SGD based on equation (9.19), we find τ and $\mathbb{E}[\mathrm{x}^\mathsf{T}\mathrm{x}]$.

From the last few sentences, $\tau = 10$. For any x, it will belong to exactly 8 stripe tilings. And, it will belong to exactly 6 pairwise interactions so 12 tilings. Thus, in total it will have 20 non-zero entries that equal to 1 and therefore $\mathbb{E}[\mathrm{x}^\mathsf{T}\mathrm{x}] = 20$.

Hence, by equation (9.19), $\alpha \doteq (10 * 20)^{-1} = \frac{1}{200}$

■

PS: Yes you are right, book has no more practice problems for this Chapter. Especially no questions for the ANN part. I will try to add a few more with later updates.