

## Dependable Learning System Fall 2018 Homework 3

**Chun-Hao Huang UIN:627002917**

This report presents the results of linear gradient-descent  $Q(\lambda)$  with binary features.

### 1. H=9, K=9

$Q(s, a)$  values for all (s, a) pairs

state \ action	a1	a2
s1	0.4282	0.2481
s2	0.5012	0.2472
s3	0.4341	0.1810

The number of visits for each (s, a) pair

state \ action	a1	a2
s1	33	1
s2	12	2
s3	31	2

### 2. H=27, K=27

$Q(s, a)$  values for all (s, a) pairs

state \ action	a1	a2
s1	0.4599	0.2665
s2	0.5141	0.2525
s3	0.4481	0.1940

The number of visits for each (s, a) pair

state \ action	a1	a2
s1	261	18
s2	116	5
s3	310	19

The results from these two settings are convergent. It can get similar results if we set H and K to greater values. Since the results for  $Q(s, a)$  depend on random number generation, I set the random number with seed 10 so that I could reproduce the same results. It would get the different results if changing the seed number, but the discrepancy is not too great. As of the number of visits for each (s, a) pair, I use an array to record if certain (s, a) is visited after choosing the action. If we observed the results for the number of visits for each (s, a) pair, it would find a trend that the number of times to take action a1 is greater than the number of times to take action a2. This trend is associated with the features we set. In addition, Q value with action a1 is also greater than Q value with action a2. The reason is that action a1 is taken much frequently.