**Programming Assignment 3**

**Reinforcement Learning**

**Jayalakshmi Sureshkumar, Eihal Alowaisheq, Tan-Hsun Weng**

1. **Value iteration**
2. **Reward = -1**

**Values:**

|  |  |  |  |
| --- | --- | --- | --- |
| 36.0676722892 | 43.8704623111 | 63.6675485404 | 76.6434240617 |
| 30.4496634735 | -5.38984798888 | 52.6904623111 | 63.6675485404 |
| 30.2542045228 | 37.0723285951 | 43.6101520111 | 52.6904623111 |
| 25.3451551908 | XXXXXXXX | 37.0723285951 | XXXXXXXX |
| 21.0347704114 | 25.3451551908 | 30.2542045228 | 26.2287840706 |

**Policies:**

|  |  |  |  |
| --- | --- | --- | --- |
| Right | Right | Right | Up |
| Up | Right | Right | Up |
| Right | Right | Up | Up |
| Up | X | Up | X |
| Up | Right | Up | Left |

1. **Reward = 0**

**Values:**

|  |  |  |  |
| --- | --- | --- | --- |
| 41.0400939488 | 47.9913293737 | 66.9704986731 | 78.766749147 |
| 36.0352044428 | -1.26349817171 | 56.9913293737 | 66.9704986731 |
| 36.5947313844 | 42.7930259956 | 48.7365018283 | 56.9913293737 |
| 32.1319592643 | XXXXXXXX | 42.7930259956 | XXXXXXXX |
| 28.2134276467 | 32.1319592643 | 36.5947313844 | 32.935258246 |

**Policies:**

|  |  |  |  |
| --- | --- | --- | --- |
| Right | Right | Right | Up |
| Up | Right | Right | Up |
| Right | Right | Right | Up |
| Up | X | Up | X |
| Up | Right | Up | Left |

1. **Q Value Learning**
2. **Reward = -1**

|  |  |  |  |
| --- | --- | --- | --- |
| down:49.1 up:57.3 left:57.3 right:65.3 | down:11.8 up:63.5 left:57.3 right:74.2 | down:63.2 up:72.3 left:65.3 right:84.0 | down:82.3 up:93.8 left:85.2 right:95.0 |
| down:41.3 up:56.4 left:49.4 right:12.1 | down:-2.5 up:12.4 left:0.1 right:13.3 | down:53.6 up:67.2 left:11.8 right:71.0 | down:61.0 up:82.2 left:62.7 right:71.2 |
| down:35.0 up:45.2 left:41.2 right:45.7 | down:47.1 up:20.6 left:41.5 right:52.0 | down:46.0 up:56.8 left:47.0 right:58.4 | down:60.3 up:66.9 left:52.8 right:60.0 |
| down:29.7 up:37.7 left:35.3 right:35.2 | X | down:37.5 up:50.7 left:45.6 right:46.0 | X |
| down:29.7 up:32.5 left:29.8 right:32.8 | down:33.3 up:32.5 left:29.7 right:36.0 | down:37.5 up:40.8 left:33.4 right:33.0 | down:33.0 up:34.4 left:36.1 right:32.7 |

1. **Reward = 0**

|  |  |  |  |
| --- | --- | --- | --- |
| down:53.7 up:60.8 left:60.8 right:68.0 | down:14.5 up:67.3 left:60.8 right:76.0 | down:65.6 up:74.4 left:67.5 right:85.0 | down:83.9 up:93.9 left:86.0 right:95.0 |
| down:46.7 up:60.0 left:53.2 right:14.6 | down:0.8 up:15.1 left:3.1 right:17.0 | down:55.8 up:55.9 left:14.6 right:74.5 | down:62.2 up:83.4 left:65.2 right:71.9 |
| down:40.8 up:52.5 left:46.9 right:50.1 | down:50.6 up:23.5 left:47.0 right:57.3 | down:49.3 up:60.8 left:50.1 right:63.4 | down:63.2 up:69.1 left:56.3 right:61.9 |
| down:36.1 up:45.7 left:41.3 right:40.6 | X | down:42.4 up:55.8 left:50.1 right:49.1 | X |
| down:35.6 up:40.5 left:36.0 right:37.8 | down:37.6 up:37.6 left:36.2 right:42.8 | down:42.1 up:48.8 left:37.9 right:38.0 | down:37.7 up:38.0 left:42.8 right:37.9 |