2D grid-world: Starting from blue reach green

1 - Data collection: Use random policy to generate data

cell numbers: agent visitation

0	0	0	0	0
0	0	0	0	0
0	0	X	0	0
0	0	0	0	0
1	0	0	0	0

2D grid-world: Starting from blue reach green

1 - Data collection: Use random policy to generate data
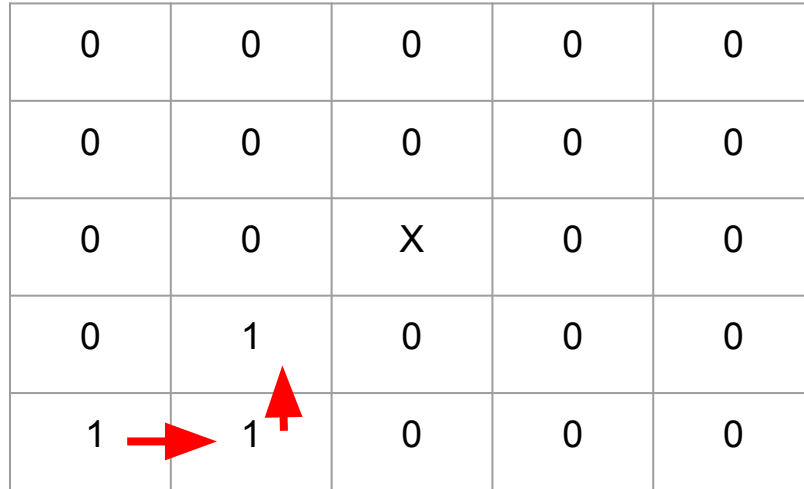
cell numbers: agent visitation

0	0	0	0	0
0	0	0	0	0
0	0	X	0	0
0	0	0	0	0
1	1	0	0	0

2D grid-world: Starting from blue reach green

1 - Data collection

cell numbers: agent visitation

0	0	0	0	0
0	0	0	0	0
0	0	X	0	0
0	1	0	0	0
1	1	0	0	0

2D grid-world: Starting from blue reach green

1 - Data collection: Use random policy to generate data

cell numbers: agent visitation
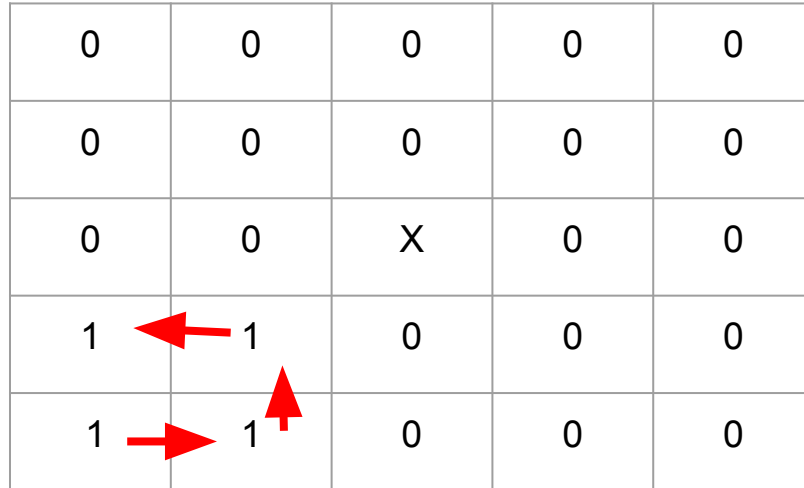
0	0	0	0	0
0	0	0	0	0
0	0	X	0	0
1	1	0	0	0
1	1	0	0	0

2D grid-world: Starting from blue reach green

1 - Data collection: Use random policy to generate data

cell numbers: agent visitation

0	0	0	0	0
0	0	0	0	0
0	0	X	0	0
1	2	0	0	0
1	1	0	0	0

After many iterations ……All cells should be visited uniformly.

2D grid-world: Starting from blue reach green

1 - Data collection: Use random policy to generate data

cell numbers: agent visitation

N	N	N	N	N
N	N	N	N	N
N	N	X	N	N
N	N	N	N	N
N	N	N	N	N

Show here environment and distribution data!!!

2D grid-world: Starting from blue reach green

Collected state-action data distribution

2D grid-world: Starting from blue reach green

2 - Optimal policy

cell numbers: agent visitation

0	0	0	1 →	1
0	1 →	1 →	1 ↑	0
0	1 ↑	X	0	0
0 →	1 ↑	0	0	0
1 ↑	0	0	0	0

2D grid-world: Starting from blue reach green

Optimal policy state-action distribution: