**گزارش بخش دوم پروژه نهایی هوش مصنوعی**

حامد تابعی9912762198

مدنی خواه9722762155

جاودانی4002134

خروجی سوال اول

Gamma: 1

Optimal policy:

[0. 0. 0. 0. 0. 0. 1. 0.]

State values:

{0: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 1: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 2: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 3: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 4: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 5: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 6: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 7: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}}

Gamma: 0.9

Optimal policy:

[0. 0.59049 0.6561 0.729 0.81 0.9 1. 0. ]

State values:

{0: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 1: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 2: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 3: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 4: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 5: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 6: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 7: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}}

Gamma: 0.5

Optimal policy:

[0. 0.03125 0.0625 0.125 0.25 0.5 1. 0. ]

State values:

{0: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 1: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 2: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 3: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 4: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 5: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 6: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 7: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}}

Gamma: 0.1

Optimal policy:

[0.e+00 1.e-05 1.e-04 1.e-03 1.e-02 1.e-01 1.e+00 0.e+00]

State values:

{0: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 1: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 2: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 3: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 4: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 5: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 6: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 7: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}}

**تحلیل تغییرات بر اساس تغییرات گاما در سوال اول**

در این خروجی، برای هر یک از مقادیر مختلف gamma، سه چیز نمایش داده شده است: مقدار gamma، سیاست بهینه و ارزش هر حالت. gamma نشان دهنده اهمیت پاداش‌های آینده در تصمیم‌گیری است. با افزایش مقدار gamma، اهمیت پاداش‌های آینده بیشتر می‌شود و در نتیجه، عامل تلاش می‌کند بیشتر به دنبال پاداش‌های بلندمدت باشد. این باعث می‌شود که عامل در تلاش برای رسیدن به هدف، بیشتر در نظر گرفتن پاداش‌های آینده باشد.

با توجه به خروجی، می‌توان دید که با افزایش مقدار gamma، سیاست بهینه و ارزش حالت‌ها بهبود یافته است. در واقع، با افزایش مقدار gamma، عامل بیشتر به دنبال پاداش‌های بلندمدت و کلی تر است و در نتیجه، سیاست بهینه و ارزش حالت‌ها بهبود می‌یابد.

با افزایش مقدار gamma، سیاست بهینه تغییر کرده است و در نتیجه، ارزش حالت‌ها نیز تغییر کرده است. برای مقدار gamma برابر 1، سیاست بهینه شامل حرکت به سمت هدف است و ارزش حالت‌ها نیز برای حالات موجود در مسیر به هدف برابر 1 و برای سایر حالات برابر 0 است. با کاهش مقدار gamma، عامل کمتر به دنبال پاداش‌های آینده است و در نتیجه، سیاست بهینه شامل حرکت به سمت هدف و تلاش برای جلوگیری از رسیدن به چاله‌ها است. همچنین، با کاهش مقدار gamma، ارزش حالت‌ها نیز تغییر کرده و برای حالات موجود در مسیر به هدف بالاتر از سایر حالات است.

بنابراین، با توجه به خروجی، می‌توان گفت که تغییرات در مقدار gamma، تاثیر قابل توجهی بر روی سیاست بهینه و ارزش حالت‌ها دارد.

خروجی سوال دوم

Gamma: 1

Optimal policy:

[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 1. 0. 0. 0. 1.

0.]

State values:

{0: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 1: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 2: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 3: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 4: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 5: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 6: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 7: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 8: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 9: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 10: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 11: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 12: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 13: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 14: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 15: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 16: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 17: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 18: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 19: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 20: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 21: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 22: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 23: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 24: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}}

Gamma: 0.9

Optimal policy:

[0.4782969 0.531441 0.59049 0.6561 0.59049 0. 0.

0. 0.729 0.6561 0.59049 0.6561 0.729 0.81

0. 0.6561 0.729 0.81 0.9 1. 0.729

0.81 0.9 1. 0. ]

State values:

{0: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 1: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 2: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 3: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 4: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 5: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 6: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 7: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 8: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 9: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 10: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 11: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 12: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 13: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 14: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 15: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 16: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 17: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 18: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 19: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 20: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 21: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 22: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 23: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 24: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}}

Gamma: 0.5

Optimal policy:

[0.0078125 0.015625 0.03125 0.0625 0.03125 0. 0.

0. 0.125 0.0625 0.03125 0.0625 0.125 0.25

0. 0.0625 0.125 0.25 0.5 1. 0.125

0.25 0.5 1. 0. ]

State values:

{0: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 1: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 2: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 3: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 4: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 5: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 6: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 7: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 8: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 9: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 10: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 11: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 12: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 13: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 14: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 15: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 16: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 17: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 18: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 19: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 20: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 21: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 22: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 23: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 24: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}}

Gamma: 0.1

Optimal policy:

[1.e-07 1.e-06 1.e-05 1.e-04 1.e-05 0.e+00 0.e+00 0.e+00 1.e-03 1.e-04

1.e-05 1.e-04 1.e-03 1.e-02 0.e+00 1.e-04 1.e-03 1.e-02 1.e-01 1.e+00

1.e-03 1.e-02 1.e-01 1.e+00 0.e+00]

State values:

{0: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 1: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 2: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 3: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 4: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 5: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 6: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 7: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 8: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 9: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 10: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 11: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 12: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 13: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 14: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 15: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 16: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 17: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 18: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 19: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 20: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 21: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 22: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 23: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 24: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}}

**تحلیل سوال دوم**

با توجه به خروجی، می‌توان دید که با افزایش مقدار gamma، سیاست بهینه و ارزش حالت‌ها بهبود یافته است. در واقع، با افزایش مقدار gamma، عامل بیشتر به دنبال پاداش‌های بلندمدت و کلی تر است و در نتیجه، سیاست بهینه و ارزش حالت‌ها بهبود می‌یابد.

با افزایش مقدار gamma، سیاست بهینه تغییر کرده است و در نتیجه، ارزش حالت‌ها نیز تغییر کرده است. به عنوان مثال، در مقدار gamma برابر 1، سیاست بهینه برای رسیدن به هدف شامل حرکت به سمت راست و پایین است. همچنین، ارزش حالت‌ها نیز برای حالات نزدیک به هدف بیشتر از حالات دورتر از هدف است.

خروجی سوال سوم

Is\_slippery = false

Gamma: 0.9

Optimal policy:

[0.59049 0.6561 0.59049 0.531441 0.4782969 0. 0.729

0.6561 0.59049 0.531441 0. 0.81 0.729 0.6561

0.59049 0. 0.9 0.81 0.729 0.6561 0.

1. 0.9 0.81 0.729 ]

State values:

{0: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 1: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 2: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 3: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 4: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 5: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 6: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 7: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 8: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 9: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 10: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 11: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 12: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 13: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 14: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 15: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 16: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 17: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 18: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 19: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 20: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 21: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 22: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 23: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 24: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}}

**به ازای is\_slippery=true**

Gamma: 0.9

Optimal policy:

[0.02522277 0.04975321 0.04200012 0.03579819 0.02861879 0.

0.08558209 0.09021244 0.07727611 0.05956689 0. 0.14527797

0.13779099 0.10775663 0.07544883 0. 0.2608817 0.20624735

0.1441145 0.08569797 0. 0.51807967 0.28248754 0.1663748

0.07562183]

State values:

{0: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 1: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 2: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 3: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 4: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 5: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 6: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 7: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 8: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 9: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 10: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 11: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 12: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 13: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 14: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 15: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 16: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 17: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 18: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 19: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 20: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 21: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 22: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 23: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 24: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}}

**توضیحات سوال سوم**

دلیل این تفاوت، مقدار `is\_slippery` است. وقتی `is\_slippery` برابر با `False` است، یعنی محیط با احتمال 1 به حالت مورد نظر تغییر می‌کند. در این حالت، عامل با استفاده از سیاست بهینه، به صورت مستقیم به حالت هدف می‌رسد و در نتیجه، ارزش حالت‌ها نیز برای همه حالت‌ها صفر یا یک است.

اما وقتی `is\_slippery` برابر با `True` است، یعنی عامل با احتمالات مختلف به حالت‌های مختلف تغییر می‌کند. در این حالت، باید احتمالات رسیدن به هر یک از حالت‌ها را در نظر گرفت. در نتیجه، سیاست بهینه تغییر می‌کند و ارزش حالت‌ها نیز برای هر حالت با توجه به احتمال رسیدن به آن حالت و پاداش‌های آینده تعیین می‌شود.

بنابراین، تفاوت سیاست بهینه و ارزش حالت‌ها در دو حالت مختلف، به دلیل تفاوت در شرایط محیط است. در شرایط مختلف، عامل باید با استفاده از الگوریتم‌های گوناگون، سیاست بهینه و ارزش حالت‌های جدید را محاسبه کند.

**خروجی سوال چهارم**

Gamma: 0.9

Optimal policy:

[0.0133436 0.00576646 0.00447781 0.00699169 0.01442604 0.0315556

0.02393071 0.04447539 0. 0. 0. 0.

0.0569017 0.04818331 0.13089483 0.29113503 0.28714536 0.2269812

0.16303876 0.10987472 0.06662675 0. 0.55237528 0.43899242

0.30637122 0.20655849 0.13654492 0.0609515 0.33333333 0.

0.60454038 0.35525553 0.21910747 0.13870443 0.07786581]

State values:

{0: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 1: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 2: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 3: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 4: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 5: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 6: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 7: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 8: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 9: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 10: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 11: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 12: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 13: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 14: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 15: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 16: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 17: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 18: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 19: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 20: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 21: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 22: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 23: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 24: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 25: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 26: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 27: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 28: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 29: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 30: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 31: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 32: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 33: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 34: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}}

**توضیحات سوال چهارم**

در اینجا نیز، با استفاده از تابع policy\_iteration ، به دنبال سیاست بهینه و ارزش حالت‌ها برای محیط FrozenLake-v1 با نقشه سفارشی custom\_map\_4 هستیم. در این حالت، مقدار is\_slippery برابر با True است.

در مورد تفاوت نتایج حاصل از پاداش‌های مختلف، باید گفت که در هر دو حالت، سیاست بهینه بدست آمده استخراج شده است. با این حال، با توجه به مقدار is\_slippery، احتمال رسیدن به هر یک از حالت‌ها در آینده ممکن است متفاوت باشد. در نتیجه، ارزش حالت‌ها و سیاست بهینه نیز ممکن است در هر دو حالت متفاوت باشد.

به طور خلاصه، تفاوت نتایج حاصل از هر پاداش به دلیل تفاوت در احتمال رسیدن به حالت‌های مختلف در آینده است.

**خروجی سوال پنجم**

move\_reward: -4

Optimal policy:

[0. 0.00390972 0.01002491 0.0257049 0.06591 0.169

0.43333333 0. ]

State values:

{0: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 1: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 2: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 3: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 4: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 5: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 6: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 7: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}}

move\_reward: -2

Optimal policy:

[0. 0.00390972 0.01002491 0.0257049 0.06591 0.169

0.43333333 0. ]

State values:

{0: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 1: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 2: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 3: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 4: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 5: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 6: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 7: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}}

move\_reward: 0

Optimal policy:

[0. 0.00390972 0.01002491 0.0257049 0.06591 0.169

0.43333333 0. ]

State values:

{0: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 1: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 2: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 3: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 4: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 5: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 6: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 7: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}}

move\_reward: 2

Optimal policy:

[0. 0.00390972 0.01002491 0.0257049 0.06591 0.169

0.43333333 0. ]

State values:

{0: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 1: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 2: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 3: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 4: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 5: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 6: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 7: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}}

**توضیحات سوال پنجم**

در این مثال، همانطور که مشاهده می‌شود، با تغییر پاداش حرکت (move\_reward)، سیاست به دست آمده و ارزش هر حالت تغییر نکرده است. این به این دلیل است که تغییر پاداش حرکت بر روی سیاست و ارزش هر حالت تاثیر ندارد. به عبارت دیگر، پاداش حرکت (move\_reward) در این بازی تاثیری در به دست آوردن بهترین سیاست و ارزش هر حالت ندارد.

**خروجی سوال ششم**

move\_reward: -4

Optimal policy:

[0. 0.04343518 0.07175592 0.11854279 0.19717164 0.33361003

0.5861718 0. 0. 0.03961687 0.06544713 0.10678447

0.17195839 0.26923858 0.39189353 0.45090139 0. 0.02308986

0.03734944 0.05905105 0.09005241 0.12821633 0.15814918 0.13527042]

State values:

{0: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 1: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 2: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 3: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 4: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 5: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 6: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 7: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 8: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 9: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 10: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 11: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 12: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 13: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 14: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 15: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 16: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 17: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 18: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 19: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 20: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 21: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 22: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 23: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}}

move\_reward: -2

Optimal policy:

[0. 0.04343518 0.07175592 0.11854279 0.19717164 0.33361003

0.5861718 0. 0. 0.03961687 0.06544713 0.10678447

0.17195839 0.26923858 0.39189353 0.45090139 0. 0.02308986

0.03734944 0.05905105 0.09005241 0.12821633 0.15814918 0.13527042]

State values:

{0: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 1: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 2: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 3: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 4: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 5: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 6: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 7: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 8: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 9: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 10: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 11: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 12: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 13: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 14: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 15: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 16: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 17: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 18: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 19: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 20: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 21: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 22: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 23: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}}

move\_reward: 0

Optimal policy:

[0. 0.04343518 0.07175592 0.11854279 0.19717164 0.33361003

0.5861718 0. 0. 0.03961687 0.06544713 0.10678447

0.17195839 0.26923858 0.39189353 0.45090139 0. 0.02308986

0.03734944 0.05905105 0.09005241 0.12821633 0.15814918 0.13527042]

State values:

{0: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 1: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 2: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 3: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 4: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 5: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 6: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 7: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 8: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 9: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 10: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 11: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 12: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 13: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 14: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 15: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 16: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 17: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 18: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 19: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 20: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 21: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 22: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 23: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}}

move\_reward: 2

Optimal policy:

[0. 0.04343518 0.07175592 0.11854279 0.19717164 0.33361003

0.5861718 0. 0. 0.03961687 0.06544713 0.10678447

0.17195839 0.26923858 0.39189353 0.45090139 0. 0.02308986

0.03734944 0.05905105 0.09005241 0.12821633 0.15814918 0.13527042]

State values:

{0: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 1: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 2: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 3: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 4: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 5: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 6: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 7: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 8: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 9: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 10: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 11: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 12: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 13: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 14: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 15: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 16: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 17: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 18: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 19: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 20: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 21: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 22: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 23: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}}

**خروجی سوال هفتم**

move\_reward: -1

Optimal policy:

[0.00997829 0.01992096 0.02702612 0.02270539 0.00681162 0.01324275

0.02391081 0.04644922 0.04858659 0. 0. 0.

0.07906549 0.09273618 0.03616711 0. 0.05617936 0.12424869

0.18142763 0. 0.01888885 0.0629423 0.15360677 0.38776162

0. ]

State values:

{0: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 1: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 2: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 3: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 4: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 5: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 6: {0: 0.0, 1: 0.0, 2: 0.0, 3: 1.0}, 7: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 8: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 9: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 10: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 11: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 12: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 13: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 14: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 15: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 16: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 17: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 18: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 19: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}, 20: {0: 0.0, 1: 1.0, 2: 0.0, 3: 0.0}, 21: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 22: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 23: {0: 0.0, 1: 0.0, 2: 1.0, 3: 0.0}, 24: {0: 1.0, 1: 0.0, 2: 0.0, 3: 0.0}}