--------------------

VI data bellow

--------------------

--------------------

runtime = 1.28 seconds

1.2810635566711426 seconds to converge VI frozen lake at gamma=1, and standard epsilon

1425 iteration to converge states values to standard epsilon

365 iteration to converge to stable policy

runtime = 0.94 seconds

0.9422519207000732 seconds to converge VI frozen lake at gamma=1, and standard epsilon until stable policy

--------------------

runtime = 0.54 seconds

0.5359499454498291 seconds to converge VI frozen lake at gamma=099, and standard epsilon

662 iteration to converge states values to standard epsilon

127 iteration to converge to stable policy

runtime = 0.82 seconds

0.8180806636810303 seconds to converge VI frozen lake at gamma=099, and standard epsilon until stable policy

--------------------

runtime = 0.22 seconds

0.21371889114379883 seconds to converge VI frozen lake at gamma=09, and standard epsilon

158 iteration to converge states values to standard epsilon

48 iteration to converge to stable policy

runtime = 0.33 seconds

0.3267860412597656 seconds to converge VI frozen lake at gamma=09, and standard epsilon until stable policy

--------------------

runtime = 0.03 seconds

0.027561187744140625 seconds to converge VI frozen lake at gamma=03, and standard epsilon

17 iteration to converge states values to standard epsilon

16 iteration to converge to stable policy

runtime = 0.04 seconds

0.0349268913269043 seconds to converge VI frozen lake at gamma=03, and standard epsilon until stable policy

--------------------

PI data bellow

--------------------

--------------------

runtime = 2.36 seconds

2.3605353832244873 seconds to converge PI frozen lake at gamma=1, and standard epsilon

9 iteration to converge states values to standard epsilon

7 iteration to converge to stable policy

runtime = 3.34 seconds

3.335667133331299 seconds to converge PI frozen lake at gamma=1, and standard epsilon until stable policy

--------------------

runtime = 2.96 seconds

2.9614198207855225 seconds to converge PI frozen lake at gamma=099, and standard epsilon

5 iteration to converge states values to standard epsilon

3 iteration to converge to stable policy

runtime = 0.18 seconds

0.17508697509765625 seconds to converge PI frozen lake at gamma=099, and standard epsilon until stable policy

--------------------

runtime = 0.15 seconds

0.1488192081451416 seconds to converge PI frozen lake at gamma=09, and standard epsilon

6 iteration to converge states values to standard epsilon

4 iteration to converge to stable policy

runtime = 0.14 seconds

0.1363050937652588 seconds to converge PI frozen lake at gamma=09, and standard epsilon until stable policy

--------------------

runtime = 0.01 seconds

0.013512372970581055 seconds to converge PI frozen lake at gamma=03, and standard epsilon

6 iteration to converge states values to standard epsilon

4 iteration to converge to stable policy

runtime = 0.01 seconds

0.01390218734741211 seconds to converge PI frozen lake at gamma=03, and standard epsilon until stable policy

runtime = 0.47 seconds

0%| | 0/300000 [00:00<?, ?it/s]C:\Users\plokhal\AppData\Local\Programs\Python\Python312\Lib\site-packages\gymnasium\utils\passive\_env\_checker.py:233: DeprecationWarning: `np.bool8` is a deprecated alias for `np.bool\_`. (Deprecated NumPy 1.24)

if not isinstance(terminated, (bool, np.bool8)):

0%| | 329/300000 [00:00<05:30, 906.77it/s]C:\Users\plokhal\Music\github\Spring\_2024\_HW4\bettermdptools\_edit\algorithms\rl.py:185: UserWarning: Episode was truncated. Bootstrapping 0 reward.

warnings.warn("Episode was truncated. Bootstrapping 0 reward.")

runtime = 242.02 seconds

11 policy difference in between Q Learning and VI, gamma=1,epsilon=1 alpha=0.1

0%| | 0/300000 [00:00<?, ?it/s]C:\Users\plokhal\Music\github\Spring\_2024\_HW4\bettermdptools\_edit\algorithms\rl.py:185: UserWarning: Episode was truncated. Bootstrapping 0 reward.

warnings.warn("Episode was truncated. Bootstrapping 0 reward.")

runtime = 454.35 seconds

11 policy difference in between Q Learning and VI, gamma=1,epsilon=05 alpha=0.1

0%| | 52/300000 [00:00<09:45, 511.89it/s]C:\Users\plokhal\Music\github\Spring\_2024\_HW4\bettermdptools\_edit\algorithms\rl.py:185: UserWarning: Episode was truncated. Bootstrapping 0 reward.

warnings.warn("Episode was truncated. Bootstrapping 0 reward.")

runtime = 484.56 seconds

17 policy difference in between Q Learning and VI, gamma=1,epsilon=045 alpha=0.1

runtime = 0.40 seconds

0%| | 96/300000 [00:00<05:14, 952.29it/s]C:\Users\plokhal\Music\github\Spring\_2024\_HW4\bettermdptools\_edit\algorithms\rl.py:185: UserWarning: Episode was truncated. Bootstrapping 0 reward.

warnings.warn("Episode was truncated. Bootstrapping 0 reward.")

runtime = 462.10 seconds

19 policy difference in between Q Learning and VI, gamma=1,epsilon=0.5 alpha=03

0%| | 0/300000 [00:00<?, ?it/s]C:\Users\plokhal\Music\github\Spring\_2024\_HW4\bettermdptools\_edit\algorithms\rl.py:185: UserWarning: Episode was truncated. Bootstrapping 0 reward.

warnings.warn("Episode was truncated. Bootstrapping 0 reward.")

runtime = 330.65 seconds

13 policy difference in between Q Learning and VI, gamma=1,epsilon=0.5 alpha=003

runtime = 0.06 seconds

0%| | 0/300000 [00:00<?, ?it/s]C:\Users\plokhal\Music\github\Spring\_2024\_HW4\bettermdptools\_edit\algorithms\rl.py:185: UserWarning: Episode was truncated. Bootstrapping 0 reward.

warnings.warn("Episode was truncated. Bootstrapping 0 reward.")

runtime = 159.61 seconds

19 policy difference in between Q Learning and VI, gamma=09,epsilon=0.5 alpha=0.1

runtime = 0.01 seconds

0%| | 0/300000 [00:00<?, ?it/s]C:\Users\plokhal\Music\github\Spring\_2024\_HW4\bettermdptools\_edit\algorithms\rl.py:185: UserWarning: Episode was truncated. Bootstrapping 0 reward.

warnings.warn("Episode was truncated. Bootstrapping 0 reward.")

runtime = 113.14 seconds

27 policy difference in between Q Learning and VI, gamma=03,epsilon=0.5 alpha=0.1

--------------------

VI data bellow

--------------------

--------------------

runtime = 0.02 seconds

0.022022008895874023 seconds to converge VI Black Jack at gamma=1, and standard epsilon

14 iteration to converge states values to standard epsilon

1 iteration to converge to stable policy

runtime = 0.00 seconds

0.0 seconds to converge VI black jack at gamma=1, and standard epsilon until stable policy

--------------------

runtime = 0.02 seconds

0.022901296615600586 seconds to converge VI Black Jack at gamma=09, and standard epsilon

14 iteration to converge states values to standard epsilon

1 iteration to converge to stable policy

runtime = 0.00 seconds

0.0 seconds to converge VI black jack at gamma=09, and standard epsilon until stable policy

--------------------

runtime = 0.02 seconds

0.021934032440185547 seconds to converge VI Black Jack at gamma=05, and standard epsilon

14 iteration to converge states values to standard epsilon

1 iteration to converge to stable policy

runtime = 0.00 seconds

0.0 seconds to converge VI black jack at gamma=05, and standard epsilon until stable policy

--------------------

runtime = 0.02 seconds

0.017939090728759766 seconds to converge VI Black Jack at gamma=01, and standard epsilon

11 iteration to converge states values to standard epsilon

1 iteration to converge to stable policy

runtime = 0.00 seconds

0.0 seconds to converge VI black jack at gamma=01, and standard epsilon until stable policy

--------------------

PI data bellow

--------------------

--------------------

runtime = 0.04 seconds

0.035890817642211914 seconds to converge PI Black Jack at gamma=1, and standard epsilon

4 iteration to converge states values to standard epsilon

2 iteration to converge to stable policy

runtime = 0.01 seconds

0.010938644409179688 seconds to converge PI black jack at gamma=1, and standard epsilon until stable policy

--------------------

runtime = 0.04 seconds

0.0393824577331543 seconds to converge PI Black Jack at gamma=09, and standard epsilon

4 iteration to converge states values to standard epsilon

2 iteration to converge to stable policy

runtime = 0.01 seconds

0.010965108871459961 seconds to converge PI black jack at gamma=09, and standard epsilon until stable policy

--------------------

runtime = 0.03 seconds

0.03189849853515625 seconds to converge PI Black Jack at gamma=05, and standard epsilon

3 iteration to converge states values to standard epsilon

1 iteration to converge to stable policy

runtime = 0.00 seconds

0.0 seconds to converge PI black jack at gamma=05, and standard epsilon until stable policy

--------------------

runtime = 0.03 seconds

0.02689981460571289 seconds to converge PI Black Jack at gamma=01, and standard epsilon

3 iteration to converge states values to standard epsilon

1 iteration to converge to stable policy

runtime = 0.00 seconds

0.0 seconds to converge PI black jack at gamma=01, and standard epsilon until stable policy

runtime = 0.02 seconds

0%| | 0/100000 [00:00<?, ?it/s]C:\Users\plokhal\AppData\Local\Programs\Python\Python312\Lib\site-packages\gymnasium\utils\passive\_env\_checker.py:233: DeprecationWarning: `np.bool8` is a deprecated alias for `np.bool\_`. (Deprecated NumPy 1.24)

if not isinstance(terminated, (bool, np.bool8)):

runtime = 10.19 seconds

24 policy difference in between Q Learning and VI, gamma=1,epsilon=1 alpha=0.1

runtime = 10.97 seconds

27 policy difference in between Q Learning and VI, gamma=1,epsilon=03 alpha=0.1

runtime = 10.72 seconds

30 policy difference in between Q Learning and VI, gamma=1,epsilon=01 alpha=0.1

runtime = 0.02 seconds

runtime = 10.92 seconds

52 policy difference in between Q Learning and VI, gamma=1,epsilon=0.3 alpha=03

runtime = 10.66 seconds

20 policy difference in between Q Learning and VI, gamma=1,epsilon=0.3 alpha=003

runtime = 0.02 seconds

runtime = 11.12 seconds

27 policy difference in between Q Learning and VI, gamma=05,epsilon=0.3 alpha=0.1

runtime = 0.01 seconds

runtime = 10.70 seconds

28 policy difference in between Q Learning and VI, gamma=01,epsilon=0.3 alpha=0.1