learning rate -> high -> more fluctuation
less fluctuation -> batch size is more
batch size is large -> learning would be slow but it would be stable
batch size vey small -> large fluactions

gamma is less -> learns faster -> but will have fluactioons
gamma is more -> learns little slow -> stable

alpha = 0.1
gamma = 0.9
iteration = 100
batch_size = 2000

Chakra

Iteration: 50 AverageReturn: -20.04 |theta|_2: 5.02
Theta [[-3.53680934e+00 2.02234294e-03 -1.17076334e-02]
[3.81642943e-03 -3.56166390e+00 5.48050205e-03]]

Iteration: 99 AverageReturn: -18.45 |theta|_2: 8.75
Theta [[-6.13491686e+00 1.09328930e-01 8.57214199e-04]
[1.08538262e-01 -6.23993337e+00 -3.92249279e-04]]

Visham
Iteration: 99 AverageReturn: -33.81 |theta|_2: 6.92
Theta [[-3.18453861 0.02870256 -0.04976909]
[0.02808518 -6.14852013 -0.05985958]]

Iteration: 50 AverageReturn: -36.28 |theta|_2: 4.65
Theta [[-1.95259924 -0.07689259 -0.04770439]
[0.00986787 -4.21552925 -0.01878029]]

alpha = 0.1
gamma = 0.9
iteration = 100
batch_size = 200

Chakra
Iteration: 50 AverageReturn: -20.15 |theta|_2: 4.65
Theta [[-3.32581369 -0.04606983 0.00348379]
[0.01509027 -3.25603638 0.0485081]]

Iteration: 99 AverageReturn: -19.21 |theta|_2: 6.96
Theta [[-4.96970963 0.04536217 -0.01116674]
[0.0882065 -4.86994977 -0.02395269]]

Iteration: 200 AverageReturn: -18.51 |theta|_2: 9.89
Theta [[-6.95875364 0.17427046 -0.03082633]
[0.2244803 -7.01602897 0.06720352]]

Visham

Iteration: 99 AverageReturn: -39.52 |theta|_2: 5.40
Theta [[-2.21076464 -0.32071899 0.02222874]
[0.07532522 -4.91103248 0.12024083]]

Iteration: 50 AverageReturn: -36.67 |theta|_2: 3.98
Theta [[-1.45015517 -0.10499996 0.20503612]
[0.01357499 -3.69607341 0.1653882]]

alpha = 0.3
gamma = 0.9
iteration = 100
batch_size = 200

Iteration: 50 AverageReturn: -18.66 |theta|_2: 8.16
Theta [[-5.93270372 0.47944897 0.06299266]
[0.04371186 -5.57320077 -0.2100371]]

Visham

Iteration: 50 AverageReturn: -32.96 |theta|_2: 5.88
Theta [[-2.35344097 -0.37706444 0.14977077]
[0.07777555 -5.37444033 0.19481875]]

Iteration: 99 AverageReturn: -36.41 |theta|_2: 8.43
Theta [[-3.55578118 -0.23138762 -0.00848821]
[-0.31721638 -7.63459996 0.21272805]]

alpha = 0.6

gamma = 0.9
iteration = 100
batch_size = 200

Iteration: 25 AverageReturn: -18.28 |theta|_2: 8.88
Theta [[-6.38323084 -0.04271546 -0.41003344]
[-0.28015218 -6.14604637 0.1970933]]

Visham

Iteration: 26 AverageReturn: -33.03 |theta|_2: 6.91
Theta [[-3.47813976 0.21747348 -0.46863363]
[-0.07637423 -5.92902781 -0.4435358]]

alpha = 0.6
gamma = 0.9
iteration = 100
batch_size = 20

Iteration: 99 AverageReturn: -20.08 |theta|_2: 12.27
Theta [[-10.505377 0.47811364 0.19010181]
[0.92632829 -6.22078742 -0.66357693]]
has fluactions

Visham

Iteration: 98 AverageReturn: -32.94 |theta|_2: 11.50
Theta [[-6.47699456 -0.9791653 -1.08193604]
[-2.28759984 -9.10460529 0.21092543]]

Iteration: 99 AverageReturn: -34.99 |theta|_2: 11.33
Theta [[-6.30851691 -1.20830543 -1.02367437]
[-2.41569365 -8.93113222 0.68968918]]

alpha = 0.2
gamma = 0.9
iteration = 100
batch_size = 20
Iteration: 99 AverageReturn: -20.44 |theta|_2: 7.41

Theta [[-5.41508084 -0.64524464 -0.17638207]
 [0.17081902 -4.96520308 0.63483928]]
less fluactions

Visham

 Iteration: 99 AverageReturn: -30.25 |theta|_2: 4.99
Theta [[-2.82891814e+00 5.84127142e-01 -7.14778594e-01]
 [-6.13816919e-01 -3.95236125e+00 -3.91046899e-03]]

alpha = 0.2
    gamma = 0.5
    iteration = 100
    batch_size = 20
 Iteration: 99 AverageReturn: -18.79 |theta|_2: 7.75
Theta [[-5.65954174 0.28235745 0.08410671]
 [0.46186732 -5.25904226 0.09350516]]

Visham

 Iteration: 99 AverageReturn: -31.02 |theta|_2: 4.91
Theta [[-2.37080156 -0.29457624 -0.77013569]
 [0.0346814 -4.20555054 -0.31152262]]

alpha = 0.2
    gamma = 0.1
    iteration = 100
    batch_size = 20
 Iteration: 98 AverageReturn: -18.08 |theta|_2: 7.95
Theta [[-5.82982071 -0.03478069 -0.11082189]
 [-0.37576297 -5.38439099 -0.04674111]]

Visham

 Iteration: 99 AverageReturn: -30.04 |theta|_2: 4.86
Theta [[-2.03603636 -0.03402458 -0.26269649]
 [0.44075487 -4.37818323 -0.25098922]]

alpha = 0.8
gamma = 0.1
iteration = 100
batch_size = 20
Iteration: 99 AverageReturn: -17.67 |theta|_2: 14.44
Theta [[-11.28043316 2.36182922 -1.00660391]
[-0.50661504 -8.57354521 0.95489929]]

Visham
large fluctutaipns

Iteration: 97 AverageReturn: -26.04 |theta|_2: 12.49
Theta [[-7.33315219 -4.06083327 4.66011409]
[-0.14733566 -7.99720439 -0.22573844]]

Iteration: 98 AverageReturn: -40.75 |theta|_2: 13.08
Theta [[-7.61597003 -4.01197486 5.10515076]
[-0.51798143 -8.40032468 0.01856678]]

Iteration: 99 AverageReturn: -36.17 |theta|_2: 13.16
Theta [[-7.56939241 -3.95525844 4.79208057]
[-0.1244132 -8.77670439 -0.47137525]]

alpha = 0.8
gamma = 0.1
iteration = 100
batch_size = 200
Iteration: 25 AverageReturn: -18.26 |theta|_2: 9.85
Theta [[-6.15102638 -0.97050087 0.45268349]
[0.04480115 -7.60849628 -0.23755153]]

Visham
stable, less flu

Iteration: 50 AverageReturn: -32.23 |theta|_2: 10.69
Theta [[-6.99441336 0.19313329 -1.37207287]
[-0.63860618 -7.88624253 0.90935462]]

alpha = 0.8
gamma = 0.99
iteration = 100
batch_size = 2

Iteration: 94 AverageReturn: -33.53 |theta|_2: 9.94
Theta [[-0.87991808 -1.27480829 -5.40051755]
 [-3.24418944 -7.50861778 0.58573079]]

 Iteration: 95 AverageReturn: -21.80 |theta|_2: 9.84
Theta [[-0.69326903 -1.62992809 -4.83246857]
 [-3.14157073 -7.72064172 0.90335898]]

Visham

 Iteration: 97 AverageReturn: -87.32 |theta|_2: 11.18
Theta [[-5.39588133 -4.82169567 4.26145584]
 [-2.58673244 -0.09528064 -6.91773004]]

 Iteration: 98 AverageReturn: -45.20 |theta|_2: 11.43
Theta [[-5.11371938 -4.84387964 4.06071989]
 [-2.56583801 -0.27995869 -7.61418895]]

 Iteration: 99 AverageReturn: -102.16 |theta|_2: 11.50
Theta [[-4.60213013 -5.08925065 3.71490436]
 [-2.84499991 -0.23314365 -7.95822111]]

alpha = 0.1
    gamma = 0.9
    iteration = 100
    batch_size = 2
 Iteration: 99 AverageReturn: -27.21 |theta|_2: 2.45
Theta [[-1.56435596 -0.19043832 -0.97295059]
 [-0.19883958 -1.58308011 -0.17384226]]

Visham
 Iteration: 97 AverageReturn: -53.76 |theta|_2: 1.69
Theta [[-0.39799963 0.15044012 0.0678816]
 [-0.58789632 -0.26193112 -1.49906038]]

 Iteration: 98 AverageReturn: -81.10 |theta|_2: 1.65
Theta [[-0.36548765 0.14708896 0.01216364]
 [-0.62673685 -0.30124548 -1.44640068]]

 Iteration: 99 AverageReturn: -43.76 |theta|_2: 1.70
Theta [[-0.4160105   0.18063506 0.08207395]

alpha = 0.7
gamma = 0.9
iteration = 100
batch_size = 50
Iteration: 98 AverageReturn: -17.84 |theta|_2: 15.12
Theta [[-11.74879372 -0.32119504 0.03159975]
[0.36166147 -9.48550489 0.48585242]]

Iteration: 99 AverageReturn: -18.43 |theta|_2: 14.95
Theta [[-11.82712623 -0.59633798 -0.04687553]
[-0.15147557 -9.11803387 0.42537564]]

Visham

Iteration: 98 AverageReturn: -31.09 |theta|_2: 9.03
Theta [[-4.17616572 -0.10068755 -1.00217901]
[0.09170128 -7.86236653 -1.1088974]]

Iteration: 99 AverageReturn: -34.33 |theta|_2: 9.43
Theta [[-4.08559232 -0.33288967 -0.87994492]
[0.50543149 -8.34003876 -1.22549077]]

alpha = 0.3
gamma = 0.1
iteration = 100
batch_size = 50

Iteration: 99 AverageReturn: -18.58 |theta|_2: 10.96
Theta [[-8.26697772 -0.36369436 0.01070406]
[-0.1864161 -7.18339724 0.1113286]]

Visham

Iteration: 97 AverageReturn: -29.74 |theta|_2: 6.16
Theta [[-2.59418511 -0.06695808 -0.05406826]
[-0.37386692 -5.56677842 0.18307537]]

Iteration: 98 AverageReturn: -38.36 |theta|_2: 6.18
Theta [[-2.63654336 -0.1005051 -0.06235933]
[-0.10945174 -5.58330198 0.31277973]]

Iteration: 99 AverageReturn: -34.77 |theta|_2: 6.29
Theta [[-2.77593753 -0.11542737 -0.23618489]
 [-0.17559554 -5.63743341 0.13160012]]

alpha = 0.3
    gamma = 0.8
    iteration = 100
    batch_size = 50

Iteration: 99 AverageReturn: -18.60 |theta|_2: 10.47
Theta [[-7.83259894 -0.18831561 -0.01205058]
 [-0.17001751 -6.94505206 0.14450576]]

Visham

Iteration: 98 AverageReturn: -29.14 |theta|_2: 7.30
Theta [[-2.86295827 -0.49317223 -0.11873376]
 [0.5942371 -6.65663331 -0.3925283]]

Iteration: 99 AverageReturn: -38.71 |theta|_2: 7.38
Theta [[-2.775508   -0.57226655 -0.34196663]
 [0.71553763 -6.75855799 -0.42660421]]