CS365 Lab D Report
Huong Phan

Overall comment:
- With higher learning rate and higher epochs, it is more likely to get stuck (yield OverflowError)
- With every epoch size tested as below except for 1000, the final loss either converges to a very small number (< 0.1 and smaller) or converges really slowly and stays at around 0.35 and above.
- Lower learning rate requires greater number of epochs to reduce loss to < 0.01.

I. Results:

1. **Learning rate = 0.3**

Note: Number of hidden nodes = 2 for all the following runs:
After about 20 iterations, I observe that final loss for **1000 epochs, 0.3 learning rate** ranges from about 0.23 to 0.60. Here are a few instances:

1000 epochs

```
Final loss: 0.6184164165586772
Final weights and biases:
{'hidden_weights': array([[ 3.25547108,  3.23475795],
       [-0.93503215, -0.94397216]]), 'hidden_biases': array([[-1.43590488],
       [ 0.7700511 ]]), 'output_weights': array([[1.57712895, 0.03433355]]), 'output_biases': array([[-
1.10454987]])}
--------
Final loss: 0.5083448447527136
Final weights and biases:
{'hidden_weights': array([[ 9.33715244,  9.33620689],
       [-2.97621603, -2.44396013]]), 'hidden_biases': array([[-5.79991425],
       [-0.74062206]]), 'output_weights': array([[3.59884244, 0.11018762]]), 'output_biases': array([[-
2.89567688]])}
--------
Final loss: 0.3871717498548475
Final weights and biases:
{'hidden_weights': array([[ 9.76958102, -6.88995851],
       [ 9.64368148,  7.33612186]]), 'hidden_biases': array([[ 0.84103686],
       [-5.48076027]]), 'output_weights': array([[-3.6719275 ,  4.35325446]]), 'output_biases': array([[-
0.67696556]])}
--------
Final loss: 0.23037648543124417
Final weights and biases:
{'hidden_weights': array([[  7.0258401 , -11.65228435],
       [ 17.81567032, -17.81586574]]), 'hidden_biases': array([[-7.36083822],
       [ 1.40421217]]), 'output_weights': array([[ 7.53214086, -3.9821244 ]]), 'output_biases':
array([[1.8044784]])}
--------
```

At learning rate = 0.3, for any epoch size 10000, there would be instances where loss converges really slowly and get stuck at > 0.35, while converges quickly to < 0.1 in other instances

10000 epochs
Loss > approximately 0.35

```
Final loss: 0.4807220650040416
```

Final weights and biases:
{'hidden_weights': array([[-20.23233398, -19.56690674],
[23.34608513, 23.34608313]]), 'hidden_biases': array([[-2.23594985],
[-17.23499747]]), 'output_weights': array([[0.03596774, 6.16547011]]), 'output_biases': array([[-5.46748945]])}

Final loss: 0.3507919709696624
Final weights and biases:
{'hidden_weights': array([[18.33365909, 23.45397501],
[-20.1832143 , 25.6633486]]), 'hidden_biases': array([[-16.67620212],
[0.87855894]]), 'output_weights': array([[7.82466262, -6.97022136]]), 'output_biases': array([[-0.85529755]])}

Loss approximately 0.01

Final loss: 0.010810425550822417
Final weights and biases:
{'hidden_weights': array([[45.66669249, 45.66669249],
[35.73707048, 36.03035648]]), 'hidden_biases': array([[-43.73721486],
[-71.84486564]]), 'output_weights': array([[10.82751679, -21.07027751]]), 'output_biases': array([[-4.89800744]])}

Final loss: 0.01049079254053916
Final weights and biases:
{'hidden_weights': array([[36.24131477, 36.45988907],
[45.90036058, 45.90036058]]), 'hidden_biases': array([[-72.77892199],
[-43.97177724]]), 'output_weights': array([[-21.22032893, 10.89697622]]), 'output_biases': array([[-4.927391]])}

20000 epochs

Loss approximately around 0.35 and above

Final loss: 0.47916224230073606
Final weights and biases:
{'hidden_weights': array([[-6.23904466, -5.38339216],
[28.29886084, 28.29886237]]), 'hidden_biases': array([[-0.9210171],
[-21.48184111]]), 'output_weights': array([[5.57807282e-03, 6.86139882e+00]]), 'output_biases': array([[-6.16537266]])}

Final loss: 0.34881501975152557
Final weights and biases:
{'hidden_weights': array([[-31.26589571, 37.24019478],
[20.44057068, 26.42777944]]), 'hidden_biases': array([[1.16258443],
[-19.02863377]]), 'output_weights': array([[-8.1298065 , 8.32634039]]), 'output_biases': array([[-0.19731451]])}

Loss approximately 0.007

Final loss: 0.007911591858321267
Final weights and biases:
{'hidden_weights': array([[125.50600133, -125.50606778],
[111.6259881 , -224.43337539]]), 'hidden_biases': array([[1.39119307],
[-111.98697341]]), 'output_weights': array([[-12.55101982, 28.96657207]]), 'output_biases': array([[5.19232015]])}

Final loss: 0.007798168015318867
Final weights and biases:

{'hidden_weights': array([[113.41456214, -226.82851502],
[125.78997286, -125.79004496]]), 'hidden_biases': array([[-113.77531561],
[1.39211702]]), 'output_weights': array([[29.04477742, -12.58561322]]), 'output_biases': array([[5.20740569]])}

50000 epochs
Loss approximately around 0.35 and above
Final loss: 0.47815501998334264
Final weights and biases:
{'hidden_weights': array([[35.26043903, 35.26043883],
[-6.55873044, -6.7375508]]), 'hidden_biases': array([[-27.51398806],
[-0.50106876]]), 'output_weights': array([[7.78415891e+00, -3.67861408e-03]]), 'output_biases': array([[-7.08960687]])}

Final loss: 0.34755653817273413
Final weights and biases:
{'hidden_weights': array([[27.743141 , 34.53621023],
[-38.19325573, 45.2778875]]), 'hidden_biases': array([[-26.1438551],
[0.90590665]]), 'output_weights': array([[9.97085843, -9.01753176]]), 'output_biases': array([[-0.95360972]])}

Loss approximately 0.004 and below
Final loss: 0.0038307125254748234
Final weights and biases:
{'hidden_weights': array([[-164.15313925, 164.1531082],
[-294.6883886 , 146.76412747]]), 'hidden_biases': array([[1.38654544],
[-147.12644181]]), 'output_weights': array([[-14.37507709, 33.45153462]]), 'output_biases': array([[5.9151395]])}

Final loss: 0.003075800577509433
Final weights and biases:
{'hidden_weights': array([[177.0959847 , -177.09600765],
[157.42456467, -315.38546388]]), 'hidden_biases': array([[1.38505134],
[-157.78729338]]), 'output_weights': array([[-14.92758139, 34.81109449]]), 'output_biases': array([[6.13374273]])}

Final loss: 0.002010792898184079
Final weights and biases:
{'hidden_weights': array([[82.43065858, 82.43065858],
[64.31805012, 63.94410922]]), 'hidden_biases': array([[-80.51623399],
[-128.34447181]]), 'output_weights': array([[14.70178283, -29.25102316]]), 'output_biases': array([[-6.56995188]])}

100000 epochs
Loss approximately around 0.35
Final loss: 0.34866769560945343
Final weights and biases:
{'hidden_weights': array([[-97.10420259, -88.44322565],
[-167.95500775, 161.57866381]]), 'hidden_biases': array([[95.64895837],
[7.84344238]]), 'output_weights': array([[9.61587592, -14.98036974]]), 'output_biases': array([[5.36427461]])}

Final loss: 0.3471044751669189
Final weights and biases:
{'hidden_weights': array([[43.33600034, 35.93893271],
       [55.37767392, -47.44525609]]), 'hidden_biases': array([[-34.20016808],
       [0.6927425]]), 'output_weights': array([[11.33585838, -9.66445808]]), 'output_biases': array([[-1.67153942]])}

Loss approximately 0.0015

Final loss: 0.001508062640082953
Final weights and biases:
{'hidden_weights': array([[-222.65522403, 222.65521357],
       [-392.04259511, 196.31749615]]), 'hidden_biases': array([[1.38240981],
       [-196.68097295]]), 'output_weights': array([[-16.71776526, 39.20034583]]), 'output_biases': array([[6.84496381]])}

Final loss: 0.0015134110995602614
Final weights and biases:
{'hidden_weights': array([[-379.08781411, 190.88776468],
       [-225.07257541, 225.07256784]]), 'hidden_biases': array([[-191.25149471],
       [1.38145139]]), 'output_weights': array([[39.19283322, -16.71133472]]), 'output_biases': array([[6.84077386]])}

Final loss: 0.0015407388923326568
Final weights and biases:
{'hidden_weights': array([[199.25972682, -399.4384114],
       [218.78889446, -218.7889083]]), 'hidden_biases': array([[-199.62291633],
       [1.3834967]]), 'output_weights': array([[39.05351994, -16.66136385]]), 'output_biases': array([[6.82425769]])}
----

2. Learning rate = 0.6

2000 epochs

Loss approximately 0.35 and above

Final loss: 0.3569046537720729
Final weights and biases:
{'hidden_weights': array([[23.27923091, -19.28077168],
       [15.026576 , 10.62722247]]), 'hidden_biases': array([[1.48845814],
       [-9.43796146]]), 'output_weights': array([[-6.21582485, 5.7375389]]), 'output_biases': array([[0.47315282]])}

Loss 0.1 and below

Final loss: 0.11872719183105514
Final weights and biases:
{'hidden_weights': array([[183.61753432, -183.65565031],
       [-184.75455868, 184.7164855]]), 'hidden_biases': array([[0.64608645],
       [0.64511711]]), 'output_weights': array([[-11.54986459, -11.5493168]]), 'output_biases': array([[13.39573411]])}

Final loss: 0.04464764880976722
Final weights and biases:
{'hidden_weights': array([[-91.04793475, 45.63044792],
       [-54.92990772, 54.92937168]]), 'hidden_biases': array([[-45.98332897],

[1.41955211]]), 'output_weights': array([[18.1088614, -8.180217]]), 'output_biases': array([[3.47580887]])}

Final loss: 0.028905513648399977
Final weights and biases:
{'hidden_weights': array([[29.6729105 , 29.6729105],
[22.53335673, 22.68608035]]), 'hidden_biases': array([[-27.71801441],
[-45.28898009]]), 'output_weights': array([[8.55731259, -16.19704638]]), 'output_biases': array([[-3.93154115]])}

10000 epochs

Loss approximately around 0.35
Final loss: 0.34897482934757884
Final weights and biases:
{'hidden_weights': array([[25.27771932, 30.91655827],
[-29.565961 , 35.98414228]]), 'hidden_biases': array([[-23.39948507],
[0.50643328]]), 'output_weights': array([[9.33713493, -7.51117484]]), 'output_biases': array([[-1.82622508]])}

Loss approximately 0.008 and below
Final loss: 0.007808534694636911
Final weights and biases:
{'hidden_weights': array([[-126.90654436, 126.90648231],
[-222.69998089, 111.120636]]), 'hidden_biases': array([[1.39054674],
[-111.48178983]]), 'output_weights': array([[-12.58506571, 29.05407596]]), 'output_biases': array([[5.20506287]])}

Final loss: 0.005098008271187697
Final weights and biases:
{'hidden_weights': array([[46.36957199, 46.67651167],
[61.12153828, 61.12153828]]), 'hidden_biases': array([[-93.12676287],
[-59.20215812]]), 'output_weights': array([[-24.74914399, 12.56095625]]), 'output_biases': array([[-5.64294807]])}

50000 epochs

Loss > 0.34
Final loss: 0.34708621724578903
Final weights and biases:
{'hidden_weights': array([[32.69557322, 40.31309001],
[-46.0396579 , 53.69376716]]), 'hidden_biases': array([[-31.26691839],
[1.10219252]]), 'output_weights': array([[10.55150027, -10.14681311]]), 'output_biases': array([[-0.4048711]])}

Loss <= 0.0015
Final loss: 0.0015071908066869202
Final weights and biases:
{'hidden_weights': array([[-221.83768843, 221.83767693],
[-395.15132444, 197.22167548]]), 'hidden_biases': array([[1.38275462],
[-197.5850614]]), 'output_weights': array([[-16.71836334, 39.19886921]]), 'output_biases': array([[6.84578576]])}

At 100000 epochs, overflow error happens often for learning rate = 0.6, so I use 80000 epochs

80000 epochs

Loss > 0.34
Final loss: 0.34690333149445224
Final weights and biases:
{'hidden_weights': array([[36.1927748 , 44.32718776],
[-52.35794673, 60.46084561]]), 'hidden_biases': array([[-34.81421394],
[1.15047314]]), 'output_weights': array([[11.0888909 , -10.83419894]]), 'output_biases': array([[-0.25481608]])}

Loss < 0.001
Final loss: 0.0009383544609889809
Final weights and biases:
{'hidden_weights': array([[-255.32246293, 255.32245671],
[-447.3568271 , 224.29689358]]), 'hidden_biases': array([[1.38124358],
[-224.66070007]]), 'output_weights': array([[-17.90851449, 42.11567861]]), 'output_biases': array([[7.31873244]])}
---

3. Learning rate = 0.1

10000 epochs
Final loss: 0.35884303063474265
Final weights and biases:
{'hidden_weights': array([[19.41120943, -15.22447695],
[15.42621769, 11.53119025]]), 'hidden_biases': array([[0.9712932],
[-9.88821541]]), 'output_weights': array([[-5.50852628, 6.06794692]]), 'output_biases': array([[-0.56146445]])}

Final loss: 0.2043756972272852
Final weights and biases:
{'hidden_weights': array([[114.14025119, -114.13992551],
[-114.57868709, 114.57836143]]), 'hidden_biases': array([[0.67353667],
[0.68206081]]), 'output_weights': array([[-9.10012026, -9.10531116]]), 'output_biases': array([[10.51910129]])}

Final loss: 0.15530484298345915
Final weights and biases:
{'hidden_weights': array([[-164.22912643, 164.22912633],
[164.54124778, -164.54124735]]), 'hidden_biases': array([[0.67784919],
[0.67784137]]), 'output_weights': array([[-10.93793682, -10.93793226]]), 'output_biases': array([[12.65737221]])}

Final loss: 0.07401316140425435
Final weights and biases:
{'hidden_weights': array([[33.44492436, -67.9437564],
[39.03779851, -39.04029134]]), 'hidden_biases': array([[-33.78307203],
[1.48257589]]), 'output_weights': array([[14.61069552, -6.8739752]]), 'output_biases': array([[3.00997279]])}

100000 epochs
Loss > 0.34
Final loss: 0.34801213139400367
Final weights and biases:
{'hidden_weights': array([[25.51423883, 31.84022659],
[-34.06222181, 40.79528775]]), 'hidden_biases': array([[-23.84183623],

[0.81628244]]), 'output_weights': array([[9.5651799 , -8.41323735]]), 'output_biases': array([[-1.15229311]])}

Loss < 0.005
Final loss: 0.0045918741690192215
Final weights and biases:
{'hidden_weights': array([[-156.02328911, 156.02326134],
[-264.81640785, 133.03808837]]), 'hidden_biases': array([[1.3855398],
[-133.40063907]]), 'output_weights': array([[-13.9233505 , 32.35658069]]), 'output_biases': array([[5.7330712]])}

Final loss: 0.004587454044615507
Final weights and biases:
{'hidden_weights': array([[155.19606286, -155.19609109],
[133.98733966, -266.75159066]]), 'hidden_biases': array([[1.38563616],
[-134.3498654]]), 'output_weights': array([[-13.9255723 , 32.36129266]]), 'output_biases': array([[5.73410026]])}

200000 epochs
Loss > 0.47
Final loss: 0.4779689091735337
Final weights and biases:
{'hidden_weights': array([[-8.40198565, -8.90780941],
[38.06488394, 38.06488423]]), 'hidden_biases': array([[-0.65026826],
[-30.01626569]]), 'output_weights': array([[-0.04530077, 8.08339387]]), 'output_biases': array([[-7.38913786]])}

Loss < 0.0015
Final loss: 0.001496307110722601
Final weights and biases:
{'hidden_weights': array([[69.34256348, 69.11435103],
[90.39007303, 90.39007303]]), 'hidden_biases': array([[-138.53968192],
[-88.47703405]]), 'output_weights': array([[-30.6807328 , 15.38190719]]), 'output_biases': array([[-6.86450322]])}

500000 epochs
Loss > 0.69
Final loss: 0.6931471805599453
Final weights and biases:
{'hidden_weights': array([[-2.68694778e-08, -8.55708650e-01],
[-1.04217964e+00, -1.61787262e+00]]), 'hidden_biases': array([[-0.23213292],
[1.33002613]]), 'output_weights': array([[-4.69827371e-08, 2.18492219e-08]]), 'output_biases': array([[5.17381649e-09]])}

Loss <= 0.0006
Final loss: 0.0005929886088252238
Final weights and biases:
{'hidden_weights': array([[92.16808707, 92.88019503],
[114.73109766, 114.73109766]]), 'hidden_biases': array([[-185.13140036],
[-112.81908992]]), 'output_weights': array([[-35.12734083, 17.50845752]]), 'output_biases': array([[-7.78945352]])}

Final loss: 0.0005907206236035743
Final weights and biases:
{'hidden_weights': array([[90.56107927, 90.5365013],
[115.82829913, 115.82829913]]), 'hidden_biases': array([[-181.1809383],

[-113.91702997]]), 'output_weights': array([[-35.15396959, 17.51831716]]), 'output_biases': array([[-7.79272546]])}

4. Learning rate = 0.01

100000 epochs

Loss > 0.35

Final loss: 0.35821764570688663
Final weights and biases:
{'hidden_weights': array([[18.94834504, -14.84302301],
[15.45969801, 11.41398128]]), 'hidden_biases': array([[1.15293845],
[-9.91300059]]), 'output_weights': array([[-5.71434368, 5.93068005]]), 'output_biases': array([[-0.21959277]])}

Final loss: 0.4868659484391682
Final weights and biases:
{'hidden_weights': array([[15.90737775, 15.90726067],
[-6.47513128, -3.68965755]]), 'hidden_biases': array([[-10.98214288],
[-0.66411154]]), 'output_weights': array([[4.97606267, -0.01075027]]), 'output_biases': array([[-4.27279813]])}

Loss <= 0.15

Final loss: 0.14122949554082934
Final weights and biases:
{'hidden_weights': array([[-183.46648134, 183.46646838],
[184.33181956, -184.33180635]]), 'hidden_biases': array([[0.67805838],
[0.67765839]]), 'output_weights': array([[-11.56314267, -11.56291245]]), 'output_biases': array([[13.38547803]])}

Final loss: 0.035358197537741345
Final weights and biases:
{'hidden_weights': array([[20.98808657, 20.47555074],
[26.76991095, 26.76991094]]), 'hidden_biases': array([[-41.52695415],
[-24.7932369]]), 'output_weights': array([[-15.11773188,   8.08807169]]), 'output_biases': array([[-3.74525286]])}

200000 epochs

Loss > 0.35

Final loss: 0.48240286852431075
Final weights and biases:
{'hidden_weights': array([[-3.33163165, -4.80490017],
[20.23418048, 20.23392736]]), 'hidden_biases': array([[-0.58865809],
[-14.55645794]]), 'output_weights': array([[-0.05697354, 5.69151293]]), 'output_biases': array([[-4.99159348]])}

Final loss: 0.3532922877402261
Final weights and biases:
{'hidden_weights': array([[25.04045912, -19.84933014],
[21.55578275, 17.08253182]]), 'hidden_biases': array([[0.58897199],
[-15.2235727]]), 'output_weights': array([[-6.14902011, 7.53760899]]), 'output_biases': array([[-1.3888393]])}

Loss < 0.03

Final loss: 0.02453051383256007
Final weights and biases:

{'hidden_weights': array([[-76.42843583,   76.42819208],
       [-130.13010203,   65.0963769]]), 'hidden_biases': array([[1.40517518],
       [-65.45334856]]), 'output_weights': array([[-9.69656077, 21.90420319]]), 'output_biases':
array([[4.06740119]])}

500000 epochs

Loss >= 0.35

Final loss: 0.4794631155271205
Final weights and biases:
{'hidden_weights': array([[26.94396054, 26.9439578],
       [-5.36678695, -6.1190531]]), 'hidden_biases': array([[-20.30037052],
       [-0.61797029]]), 'output_weights': array([[6.68585463, -0.008199]]), 'output_biases': array([[-
5.98941578]])}

Final loss: 0.3491878071947385
Final weights and biases:
{'hidden_weights': array([[19.39455561, 25.23395447],
       [-28.50681445, 34.27221186]]), 'hidden_biases': array([[-18.0215435],
       [1.21242351]]), 'output_weights': array([[8.0307351 , -7.96511195]]), 'output_biases': array([[-
0.06657821]])}

Loss >= 0.006

Final loss: 0.006266470767988731
Final weights and biases:
{'hidden_weights': array([[56.00039164, 56.00039164],
       [44.89320674, 45.34961458]]), 'hidden_biases': array([[-54.07638086],
       [-90.32203263]]), 'output_weights': array([[12.08328161, -23.72350064]]), 'output_biases': array([[-
5.43979489]])}

1000000 epochs

Loss >= 0.35

Final loss: 0.3480203385717222
Final weights and biases:
{'hidden_weights': array([[-34.62265878, 41.39132696],
       [26.39233815, 32.69142185]]), 'hidden_biases': array([[0.76383719],
       [-24.68356399]]), 'output_weights': array([[-8.36338853, 9.65389269]]), 'output_biases': array([[-
1.29083295]])}

Loss >= 0.005

Final loss: 0.004653108144900433
Final weights and biases:
{'hidden_weights': array([[153.75567956, -153.7557144],
       [136.19247931, -271.80770321]]), 'hidden_biases': array([[1.38695828],
       [-136.55465928]]), 'output_weights': array([[-13.88737757, 32.25833585]]), 'output_biases':
array([[5.72076437]])}

Final loss: 0.003040402891028484
Final weights and biases:
{'hidden_weights': array([[72.04461838, 72.04461838],
       [56.84031566, 56.95424719]]), 'hidden_biases': array([[-70.12785271],
       [-113.87610895]]), 'output_weights': array([[13.74997348, -27.24815514]]), 'output_biases': array([[-
6.15807015]])}

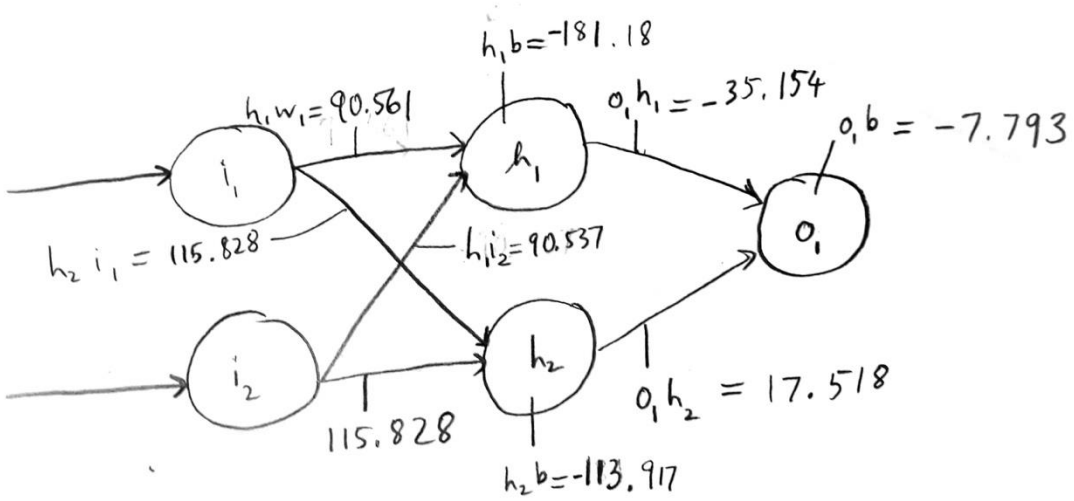
II. Further exploration

I increase the number of hidden nodes to 3, with learning rate = 0.1 and number of epochs 100000, and found that while final loss does not converge to a number >= 0.35 anymore like the cases above, sometimes it would still get stuck and produce overflow error. Also, loss for the same conditions with 3 hidden nodes generally have lower final loss than that with 2 hidden nodes, so we can assume that increasing number of hidden nodes would increase accuracy of neural networks. Here are some of the runs, all with final loss < 0.01:

```
Final loss: 0.00820528823215656
Final weights and biases:
{'hidden_weights': array([[-564.49270726,  562.63359831],
       [-176.46400759, -176.45451358],
       [ 566.01575096, -567.89907696]]), 'hidden_biases': array([[  1.31944721],
       [177.75219536],
       [  1.31003419]]), 'output_weights': array([[-23.99092285,  21.15921505, -23.95038209]]),
'output_biases': array([[12.14123608]])}
--------
Final loss: 0.005686729809009523
Final weights and biases:
{'hidden_weights': array([[-351.12347022,  360.36824943],
       [ 325.33575198, -323.27894318],
       [ 150.63541539,  150.30553011]]), 'hidden_biases': array([[  -8.02466538],
       [  -0.39288126],
       [-148.16155691]]), 'output_weights': array([[-19.4143068 , -20.38133632,  24.96130996]]),
'output_biases': array([[2.38664183]])}
--------
Final loss: 0.004975409158573285
Final weights and biases:
{'hidden_weights': array([[ 330.3705355 , -328.14250939],
       [ 100.17988517,  100.20399295],
       [-322.4713203 ,  324.59574479]]), 'hidden_biases': array([[ -0.83377579],
       [-97.96427391],
       [ -0.70101396]]), 'output_weights': array([[-20.12287968,  20.54191918, -20.18722014]]),
'output_biases': array([[6.90465063]])}
```

III. Neural network for XOR

500000 epochs, learning rate = 0.1, final loss = 0.000591


$h_1 b = -181.18$
$h_1 w_1 = 90.561$
$o_1 h_1 = -35.154$
$o_1 b = -7.793$
i_1
h_1
o_1
$h_2 i_1 = 115.828$
$h_1 i_2 = 90.537$
i_2
h_2
$o_1 h_2 = 17.518$
115.828
$h_2 b = -113.917$