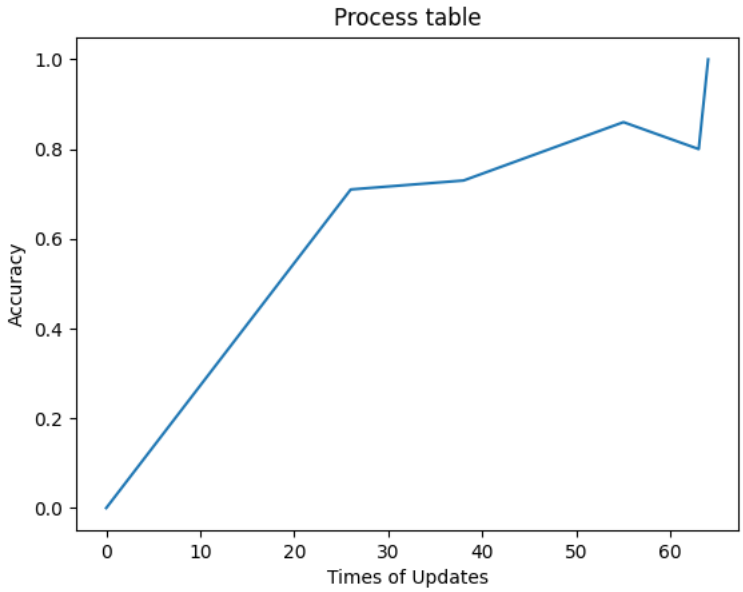
Report

All definition for the terms are listed in README.txt, except for “delta”.
Definition of Delta: If the dataset is linear separable, there exists a positive real number, delta, such that for all data vectors,
true_classification*dot_product(data_vector, weights_of_linear_separator) is greater than delta.

Therefore, I calculated every delta^2’s lower bond based on convergence theorem for dataset that is linear separable.

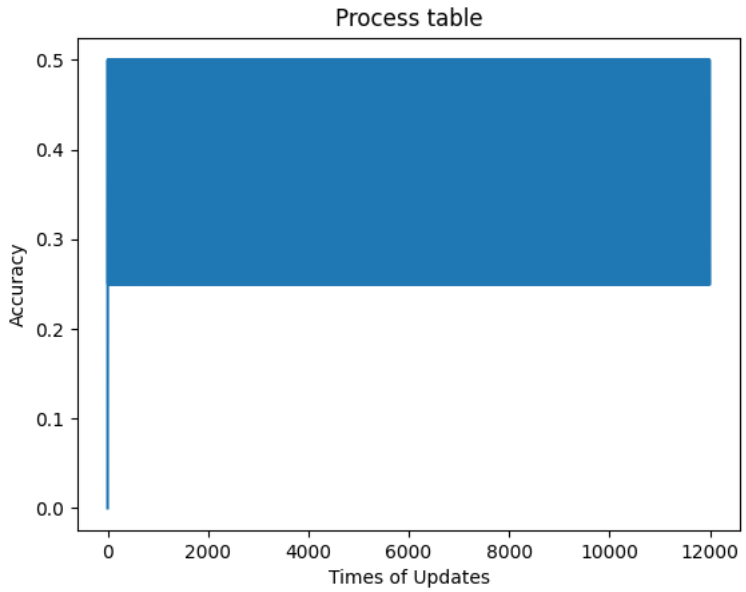
1. linearSmoke.dat:




```
chen@chen-System-Product-Name:~/Desktop/csc246_Project1$ ls
data  linearSmoke.dat  perceptronProject.pdf  perceptron_starter.py  test.py
chen@chen-System-Product-Name:~/Desktop/csc246_Project1$ python3 perceptron_starter.py --train_file data/linearSmoke.dat --iterations 2000

find it!
it uses 0.0014283670000001525sec

All data vectors are bounded by R= 1.746803700427691
it uses 64 steps(updates) and 5 iterations
Since it converges, we can calculate out delta^2 > 0.04767692449731054

Final accuracy: 1.0
Max accuracy: 1.0

Feature weights (bias last): 4.536181279855713 2.5594303280244404 -5.426706309429592 -2.0
```

Base on the 2 pictures, the accuracy quickly converges to 100% after 60 updates for weights. Therefore, it is linear sparable with delta^2 >0.047.

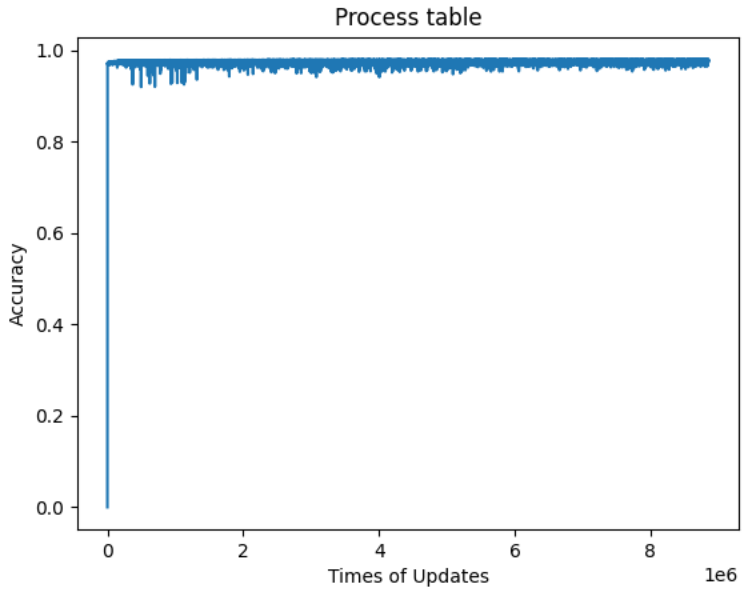
2. xorSmoke.dat:




```
chen@chen-System-Product-Name:~/Desktop/csc246_Project1$ python3 perceptron_starter.py --train_file data/xorSmoke.dat --iterations 2000

1000 iterations has been finished

Weights have been updated 10000 times

2000 iterations has been finished

cannot find it
it uses 0.046601212999999975sec

All data vectors are bounded by R= 1.7320508075688772
it uses 11998 steps(updates) and 2000 iterations

Final accuracy: 0.25
Max accuracy: 0.5

Feature weights (bias last): -2.0 -1.0 0.0
```

Based on 2 pics, the accuracy has no tendency to be increased beyond 50% even after 12000 updates. Thus, it is probably none-linear separable data set as expected

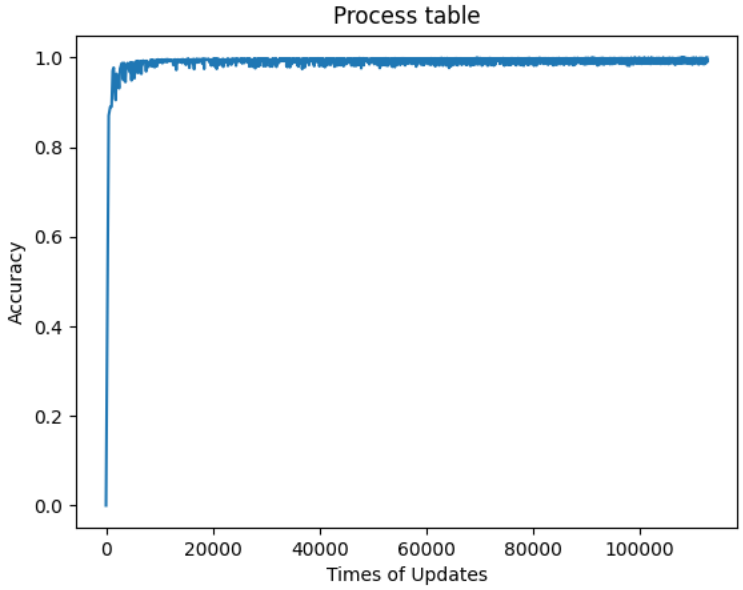
3.




```
cannot find it
it uses 649.399341626sec

All data vectors are bounded by R= 1200.4052470355523
it uses 8855802 steps(updates) and 10000 iterations

Final accuracy: 0.9769247960665997
Max accuracy: 0.980556486758297

Feature weights (bias last): 24.296875 -196.90328743378763 29558.252531670325 -3040.0567505363647 -67.20152402196032 83.8743403739924 3988.5403605128367 -344.3561343077144 -33198.0
```

I used 10000 iterations to try to find the separator, but I could not find it. Based on the table, at the tail of the plot, it does not have an evident tendency to increase further. Since the max accuracy is 98.05%, it is probably not linear separable.

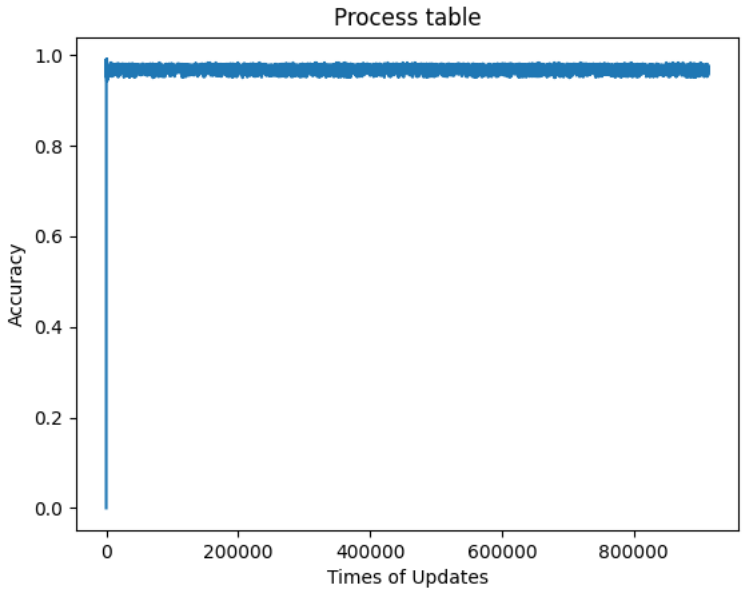
4.challenge1.dat




```
find it!
it uses 163.4135121860002sec

All data vectors are bounded by R= 4.143379680902967
it uses 112599 steps(updates) and 4174 iterations
Since it converges, we can calculate out delta^2 > 0.0001524666753711807 2

Final accuracy: 1.0
Max accuracy: 1.0

Feature weights (bias last): -5.131662805078532 -228.56419114126376 43.362348421969436 -166.356740439428 -246.26849630493854 41.151374369177 69 -72.33828212323539 171.83629895945413 -288.68588268794656 185.23478902951297 42.81688179388236 248.97477952884302 29.35045373222938
3 69.09934633386872 -225.17233100389404 -46.855160238993 28 -144.26189488159068 273.3706926718957 7 20.389012440166177 1.3344190626038936 -216.97087222568743 -50.752166675655985 191.3267958861096 -17.078012726894031 158.55817317318912 -130.9942662773926 191.9044262262456 -19.1
56278171260453 -156.01558416832248 -14.096758922836326 59.0
```

Based on the information in the terminal, the linear separator is eventually found, after 112599 updates and 4174 iterations. The lower bound of delta^2 is 0.0001

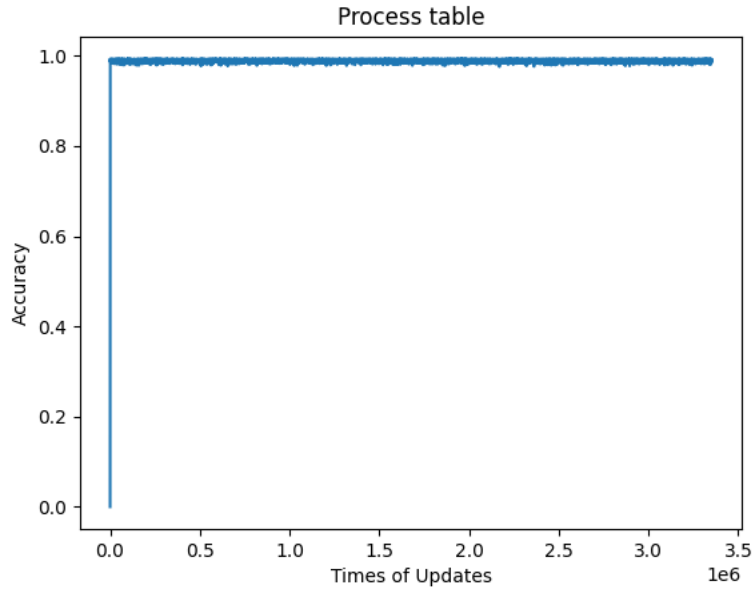
5. challenge2.dat




```
cannot find it
it uses 1984.645919031sec

All data vectors are bounded by R= 4.12563955983977
it uses 3351864 steps(updates) and 50000 iterations

Final accuracy: 0.9915
Max accuracy: 0.994

Feature weights (bias last): -1.2293268303911562 -5.648483242634707 9.55172807517887 -7.951004724368641 -0.46934084722795333 0.41048846899411154 13.587411989576875 -6.379089614401729 1.1884858659413493 3.379789697575842 6.089709020767819 -10.089172585150282 14.625460498376597 -0.4950473362549459 0.4223703131059002 3 10.281631118142505 -0.6546302526972969 3.1936430330112207 13.456003085898488 14.186096769302816 4.188473854234272 -1.9952171262810743 -2.8867098985725076 -0.10561750086257775 -4.617178595318425 -4.069163893050266 -0.5694420857994579 17.083427703585727 1.7282074194969343 1.7045568946648022 -2.0
```

I try to find the separator using 50000 iterations, but I could not find it even with 3351864 updates. As we can see, based on the table, all accuracy at the end is lower than 99.4%, since there is a peak at very beginning, and the max accuracy is recorded as 0.994. Thus, we can conclude that it does not have a tendency of increasing beyond even 0.994. Therefore, it is probably not a linear separable data set.

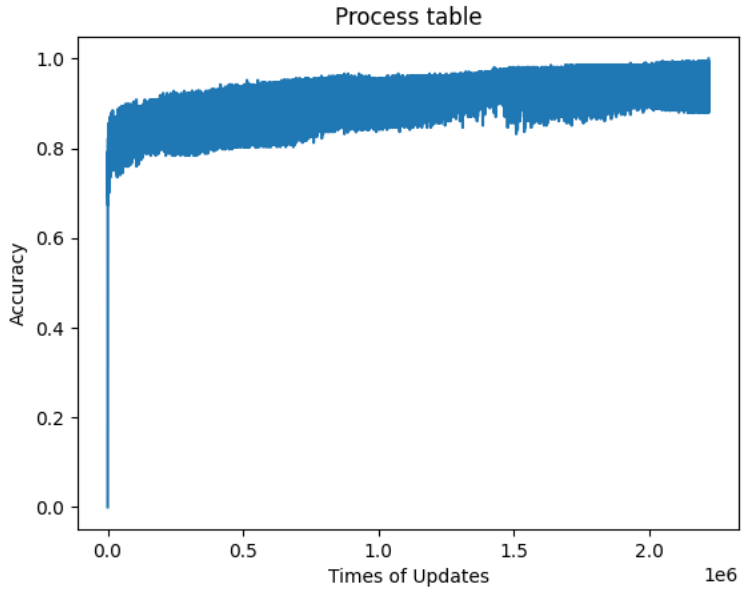
6.challenge3.dat




```
cannot find it
it uses 1984.645919031sec

All data vectors are bounded by R= 4.12563955983977
it uses 3351864 steps(updates) and 50000 iterations

Final accuracy: 0.9915
Max accuracy: 0.994

Feature weights (bias last): -1.2293268303911562 -5.648483242634787 9.55172807517887 -7.95100472436041 -0.4693408472279533 0.41040846899411154 13.587411989576875 -6.379089614401729 1.1884858659413493 3.379789697575842 6.009789028767819 -10.009172585150282 14.625460498376597 -0.4950473362549459 0.42237031310590023 10.281631118142505 -0.6546302526972969 3.1936430330112207 13.456063085898488 14.186096769382816 4.188473054234272 -1.9952171262810743 -2.8067098985725076 -0.18561750806257775 -4.617178595318425 -4.869163893050266 -0.5694420857994579 17.083427703585727 1.7282874194969343 1.7045568946648822 -2.0
```

Similar to challenge1, the accuracy is pretty stable between 0.99 and 1 exclusively. Thus, it does not have a tendency of being convergent to 1. Therefore, it is probably not a linear separable dataset.

7. challenge4.dat




```
find it!
it uses 533.447381649sec

All data vectors are bounded by R= 4.85347042421676
it uses 2219731 steps(updates) and 132312 iterations
Since it converges, we can calculate out delta^2 > 7.402078215783086e-06

Final accuracy: 1.0
Max accuracy: 1.0

Feature weights (bias last): -272.85530000088636 -90.02689999991695 607.8496999990963 -270.11769999775111 82.67130000061712 -126.39420000166805 194.65380000129971 198.42420000155988 -290.33500000142089 145.40189999959455 -123.83369999986347 -238.96790000081802 146.12780000006452 -31.219899999528142 -128.86449999682795 163.72299999999433 110.81440000007026 -150.3934000004746 115.24319999963396 -244.88039999950557 289.85120000002951 -382.97839999964642 302.77579999988887 -291.09200000019503 132.84389999993118 38.220199999762876 -113.58950000004771 50.670100000348434 53.390299999978077 -278.51970000145707 351.75289999988864 -134.47660000288258 -88.322799998723333 174.51089999780265 -133.75429999988306 17.440099999866472 165.57339999981999 -28.411700000214545 -158.75099999918118 205.57420000010197 -63.857900000168165 -74.997699999766334 -17.670999999991068 -69.649399999987698 86.313200000333237 -189.17779999992325 89.782699998880813 -379.64130000003976 -444.82509999638463 2096.528199994738 -811.21180000805743 -1346.0841000005087 -584.58029999971338 60.163000000010992 -245.59649999911372 841.54510000011838 498.63650000047157 -530.81580000219449 -344.61010000835292 120.95959999899729 161.0
```

It used 2219731 steps and 132312 iterations to find the linear separator. The lower bound of Delta^2 is shown above.