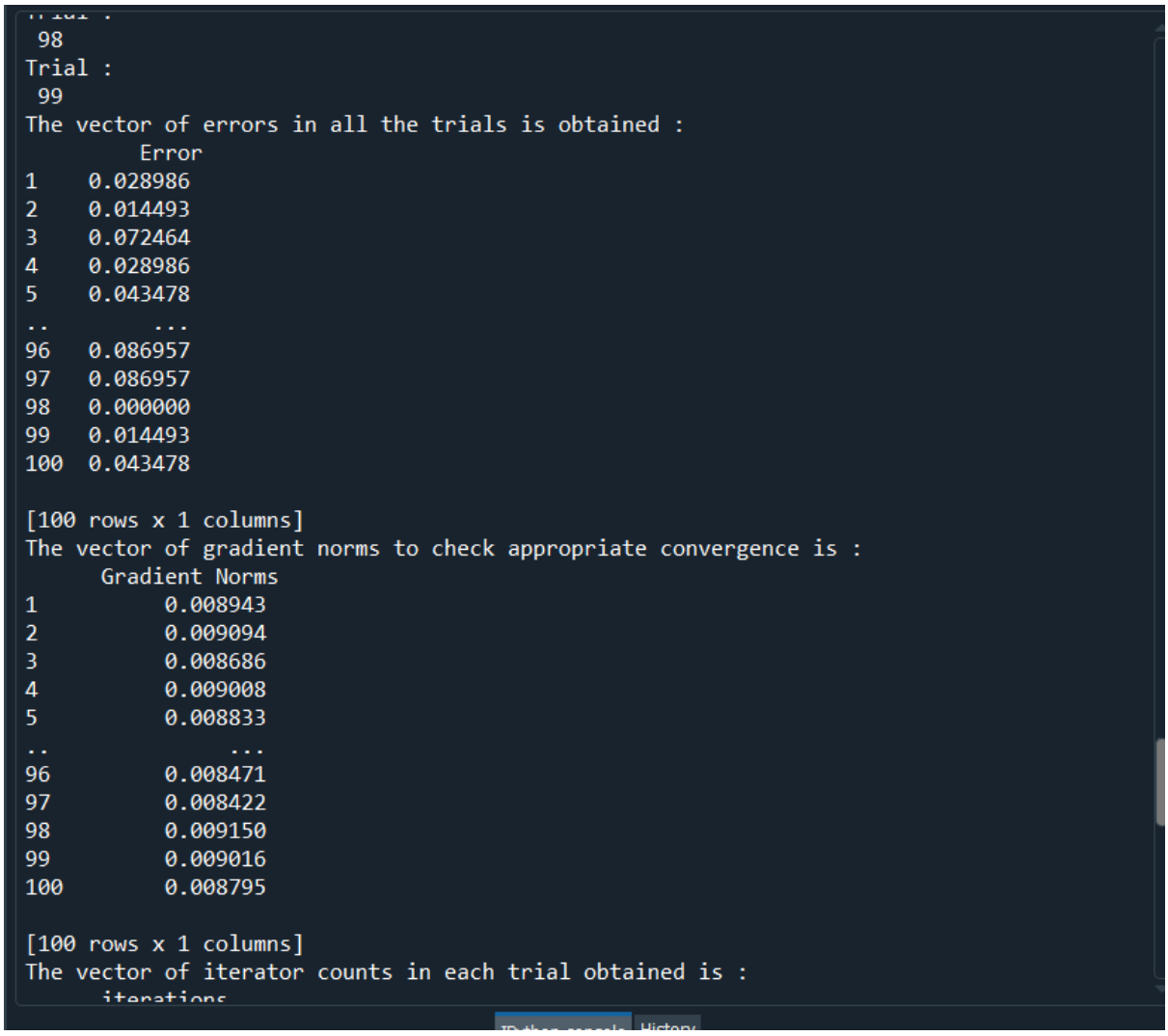
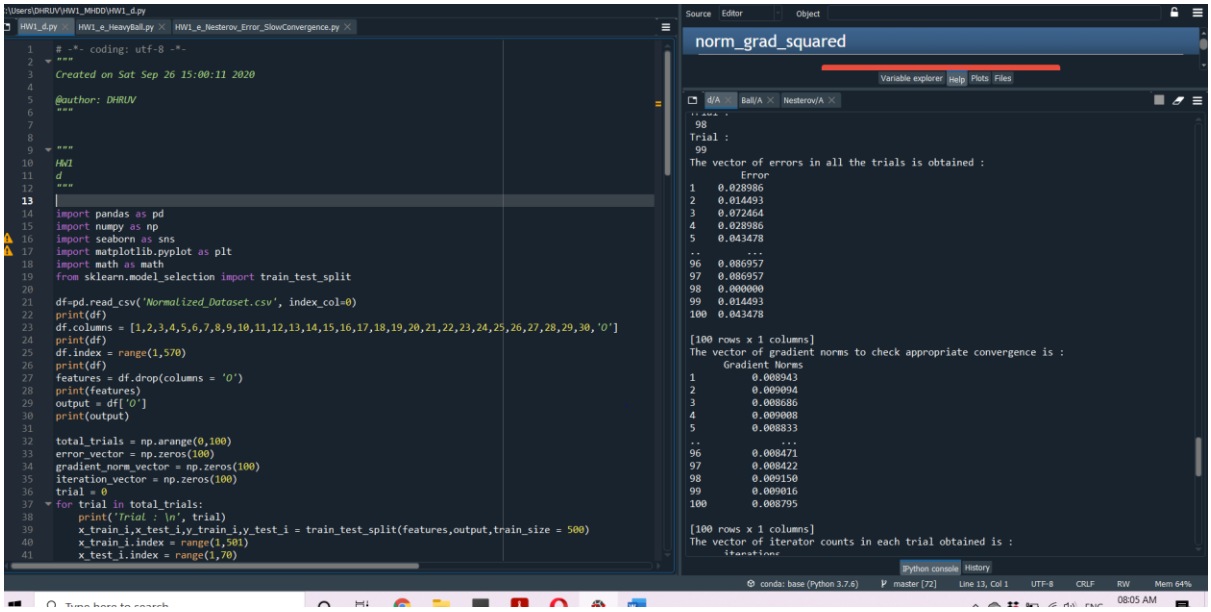
THESE DIAGRAMS ARE FOR PART d.

(It has a rather high run time)

Few Results for Gradient Descent






```
[100 rows x 1 columns]
The vector of iterator counts in each trial obtained is :
      iterations
1         5332.0
2         4651.0
3         5252.0
4         5153.0
5         4587.0
..           ...
96        4171.0
97        5191.0
98        4763.0
99        4818.0
100       4387.0

[100 rows x 1 columns]
The average iterations till the accuracy of 10^-6 is satisfied in all 100 trials is :
 4752.54
The average error obtained in all the 100 trials with 500 iterations in each trial is :
 0.05260869565217392
The average error in percentage is :
 5.2608695652173925
The average error in terms of average test data output mismatches is :
 3.630000000000001

In [2]:
```

Few Results for Heavy Ball

```
Trial :
 99
The vector of errors in all the trials is obtained :
         Error
1    0.014493
2    0.028986
3    0.086957
4    0.086957
5    0.000000
..        ...
96   0.043478
97   0.028986
98   0.072464
99   0.000000
100  0.057971

[100 rows x 1 columns]
The vector of gradient norms to check appropriate convergence is :
      Gradient Norms
1           0.009005
2           0.008738
3           0.008597
4           0.008686
5           0.009153
..               ...
96          0.008884
97          0.008880
98          0.008633
99          0.009110
100         0.008753

[100 rows x 1 columns]
The vector of iterator counts in each trial obtained is :
      iterations
```

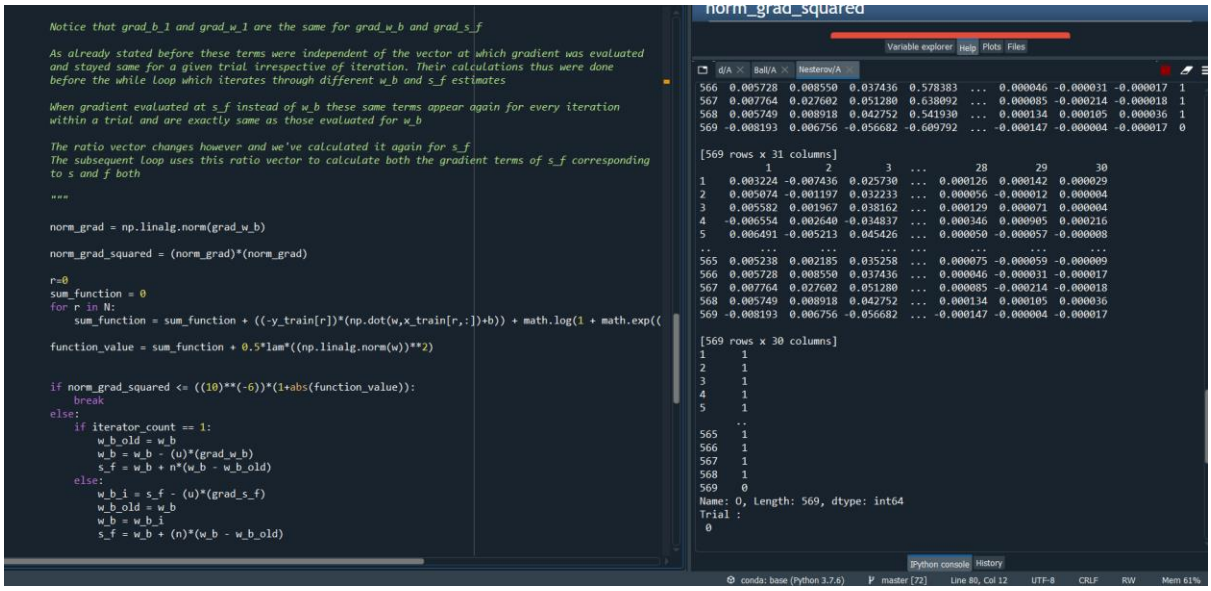
```
[100 rows x 1 columns]
The vector of iterator counts in each trial obtained is :
      iterations
1         941.0
2         850.0
3         987.0
4         879.0
5         951.0
..          ...
96        856.0
97        881.0
98        934.0
99        990.0
100       995.0

[100 rows x 1 columns]
The average iterations till the accuracy of 10^-6 is satisfied in all 100 trials is :
 931.29
The average error obtained in all the 100 trials with 500 iterations in each trial is :
 0.0518840579710145
The average error in percentage is :
 5.18840579710145
The average error in terms of average test data output mismatches is :
 3.5800000000000005

In [2]:
```

As can be seen Heavy Ball converges much faster than normal Gradient Descent

For Nesterov



The run times for Nesterov Model were rather high and so in the few trials I could observe I obtained approximate iterations of ~1000

Again it was, due to large run times, not possible for me to tune the momentum term to ensure a quick convergence

CURVES

Plots have been submitted in the folder

Which model would I select?

I would select the Heavy Ball one as it showed the best results for me, despite the Nesterov method being the most preferred one. Due to run time issues I couldn't tune the parameter of momentum in Nesterov, which weakened it's results