# Accomplishments

This week I worked on speeding up and debugging the python + GPU RbmStack, and I looked over the competitions hosted on the Kaggle web site (www.kaggle.com) to see if any of their datasets might be good candidates for my project. I also briefly reviewed the two papers and time series classification dataset links suggested by Professor Wolfe, and reviewed a competition dataset referenced by one of the papers.

## Kaggle

[**American Epilepsy Society Seizure Prediction**](http://www.kaggle.com/c/seizure-prediction)

[**Africa Soil Property Prediction Challenge**](http://www.kaggle.com/c/afsis-soil-properties)

## RbmStack / MnistAutoencoder

I made some additional progress in getting the GPU Python working like the Matlab, but something is still different between them. The table below contrasts RMSE between the two implementations for various parameterizations.

Configuration A= (784 1000 500 250 30), B= (784 2000 1000 500 30)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Epochs | Sample | Drop | Configuration | Matlab RMSE | | Python + GPU RMSE | |
|  |  |  |  | Train | Test | Train | Test |
| 10 | 0 | 0 | A | 0.1294 | 0.1275 | 0.1465 | 0.1454 |
| 10 | 0 | 1 | A | 0.2256 | 0.2261 | 0.2566 | 0.2569 |
| 10 | 1 | 0 | A | 0.1458 | 0.1450 | 0.1474 | 0.1467 |
| 10 | 1 | 1 | A | 0.2913 | 0.2939 | 0.2165 | 0.2160 |
| 10 | 0 | 0 | B | 0.1252 | 0.1232 | 0.1372 | 0.1355 |
| 10 | 0 | 1 | B | 0.2361 | 0.2364 | 0.2350 | 0.2353 |
| 10 | 1 | 0 | B | 0.1465 | 0.1456 | 0.1475 | 0.1475 |
| 10 | 1 | 1 | B | 0.3017 | 0.3043 | 0.2235 | 0.2236 |
| 50 | 0 | 0 | A | 0.1257 | 0.1235 | 0.1474 | 0.1456 |
| 50 | 0 | 1 | A | 0.2255 | 0.2252 | 0.3140 | 0.3162 |
| 50 | 1 | 0 | A | 0.1348 | 0.1344 | 0.1399 | 0.1388 |
| 50 | 1 | 1 | A | 0.2881 | 0.2898 | 0.2429 | 0.2433 |
| 50 | 0 | 0 | B | 0.1200 | 0.1175 | 0.1271 | 0.1252 |
| 50 | 0 | 1 | B | 0.2391 | 0.2388 | 0.2870 | 0.2889 |
| 50 | 1 | 0 | B | 0.1380 | 0.1387 | 0.1380 | 0.1372 |
| 50 | 1 | 1 | B | 0.2940 | 0.2958 | 0.2776 | 0.2793 |

## Data

Data sets: <http://www.cs.ucr.edu/~eamonn/time_series_data/>

These time series data sets range in length from 24 to 1882 samples and have 20 to 1800 training examples with a small number of classes. I did have not requested access to the data which requires a password to unlock the zip file.

## Papers

Review from the AI field:

* <http://www.sciencedirect.com/science/article/pii/S0952197610001727>
* Engineering Applications of Artificial Intelligence
* Volume 24, Issue 1, February 2011, Pages 164–181
* *A review on time series data mining*

Review from the forecasting field:

* <http://www.sciencedirect.com/science/article/pii/S0169207011000616#>
* International Journal of Forecasting
* Volume 27, Issue 3, July–September 2011, Pages 635–660
* *Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction*

This paper referenced the NN3 competition from which I downloaded the competition dataset. There were a total of 111 time series ranging from 62 to 162 samples in length.





