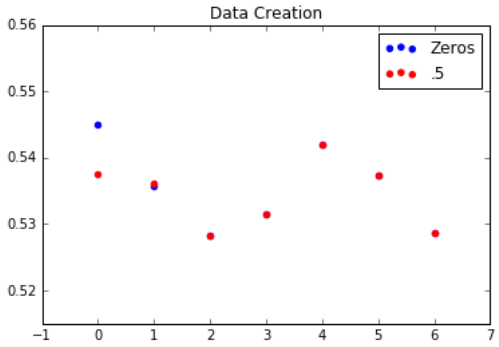
Theory of Neural Networks MP3

Michael Person

May 3, 2017

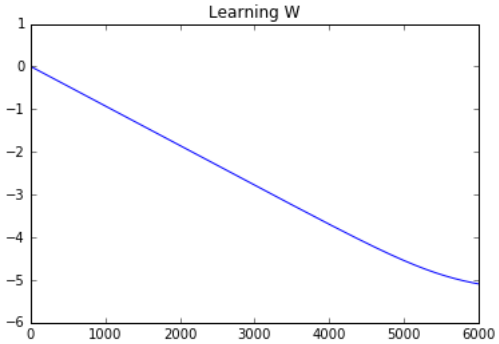
1 Task 1

1.1 Part A



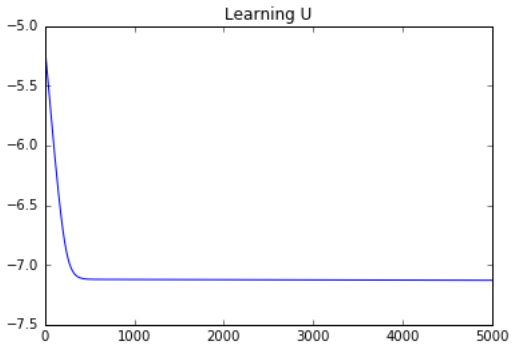
The graph of generated data is interesting because it shows that after about the 2nd or 3rd time sequence, the sample converges to the same value. This is because as the sample moves forward in time, the initial condition is less and less important therefore they both go to the same value. However before this point there is a difference in the first few time sequences.

1.2 Part B



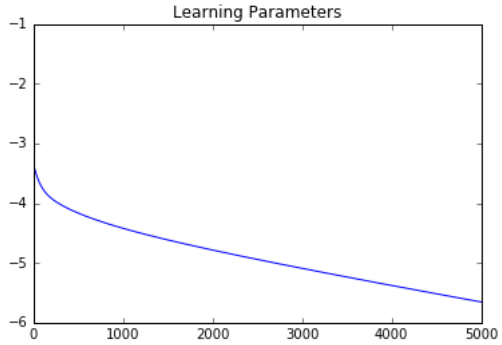
As we expected we see a monotonically decreasing MSE as the epochs increase.

1.3 Part C



What differs about learning the U matrix instead of the W matrix is that the learning has a greater effect on the MSE. This makes sense because the U matrix is much more prevalent in the network when you unravel it in time. The error quickly decreases and then levels out at a local optima that is quite good achieving errors on the order of 10^{-7}.

1.4 Part D

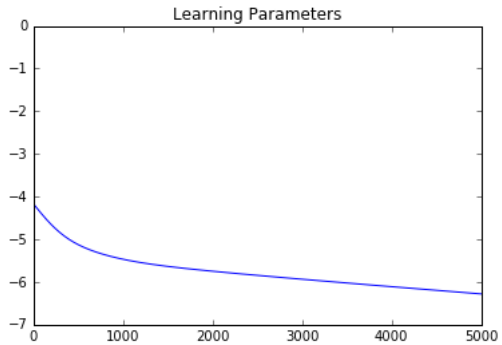


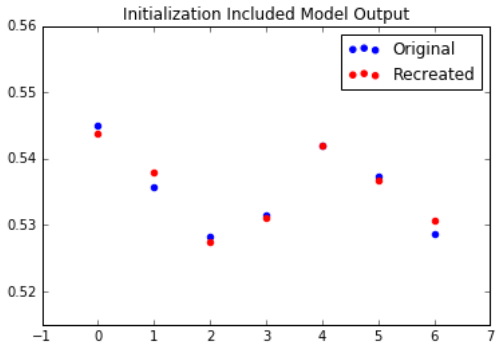
The MSE graph is monotonically decreasing which means that our learning rate isn't too large and also means that if we had let the learning continue for more than the maximum of 5000 epochs we would see a continually improving network. We don't see as fast of a convergence as when we were only optimizing U but this makes sense because we have to learn all 3 weights and not just 1 of them which will make the learning slower.

1.5 Part E

The derived gradient is:

$$\frac{\partial E}{\partial \mathbf{y}_1(0)} = \mathbf{U}^T \boldsymbol{\delta}(1) \tag{1}$$

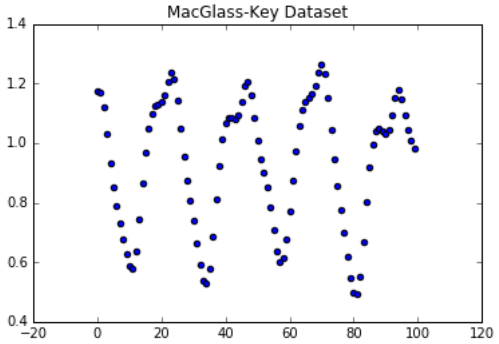




For the added learning of the initial condition, no marked improvement was found. Although the network achieves a lower MSE after 5000 epochs than when the initial condition was excluded, this is soley due to a more optimal random weight initialization. However for how cheap it is to learn the initial condition, for larger problems, it may be a worthwhile way to make the model slightly more accurate.

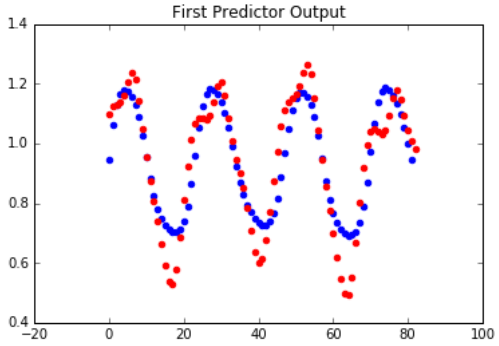
2 Task 2

2.1 Part A



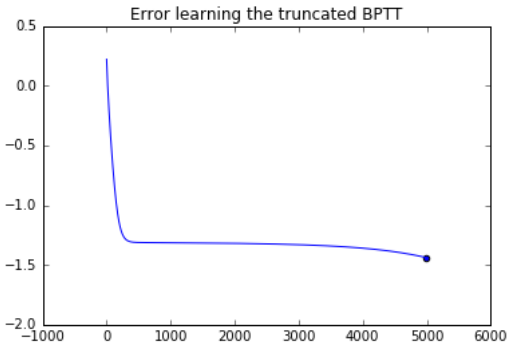
The data is highly erratic. The data seems to be largely sinusoidal however the crests of the wave are distorted. In addition, the crests do not all vary the same meaning that there is a time dependency in the distortion that occurs. These are all traits of a dataset that will most likely prove difficult to model.

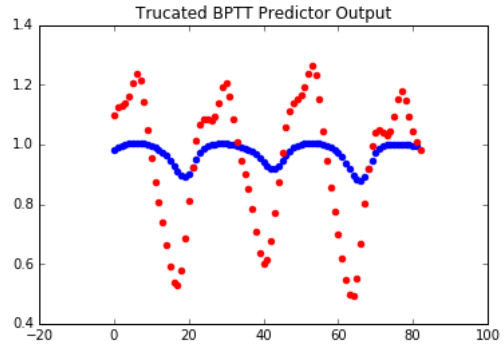
2.2 Part B



The network does a fairly good job of capturing the sinusoidal aspect of the dataset however it fails completely to capture the erratic crests and also the troughs are not low enough.

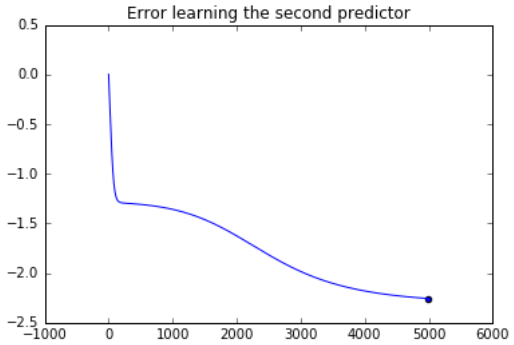
2.3 Part C

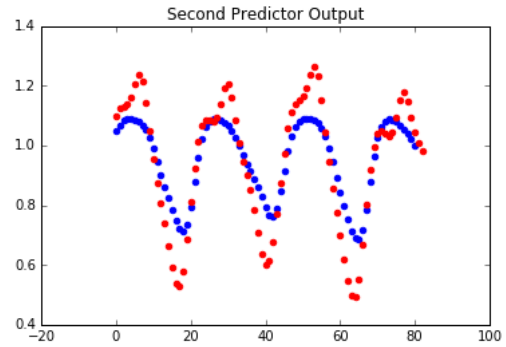




Using the truncated BPTT is far inferior to the full version. There is a small decrease in computational time however it is not that much faster in order to use it over the complete BPTT.

2.4 Part D





This prediction looks very similar to the single step ahead prediction. It even does a slightly better job of capturing some erratic behavior and not just modeling it as a sinusoidal function. However this may be due to a slightly better weight initialization. Either way it is far superior to the truncated BPTT.

2.5 Part E

The single step ahead predictor was very good and achieved an average MSE of .004 while my model achieved an average MSE of 1.03. This kind of makes sense that the point emmediately behind is relatively close to the current point and therefore the error will be small and also that we are trying to model a notoriously difficult dataset to model and are only allowing the model to train for 5000 epochs. Most likely a smaller learning rate and training for many many more epochs would be required to model the data if not a larger model.

The double step ahead predictor was again very good but not as good as the single step and achieved an average MSE of .015 while my model achieved 1.87. This is to be expected that as you move farther away from the current point, the worse your guess will be.

WORK ORIGINATION CERTIFICATION

By submitting this document, I, Michael Person, the author of this deliverable, certify that

1. I have reviewed and understood the Academic Honesty section of the current version of FITs Student Handbook available at http://www.fit.edu/studenthandbook/, which discusses academic dishonesty (plagiarism, cheating, miscellaneous misconduct, etc.)
2. The content of this Mini-Project report reflects my personal work and, in cases it is not, the source(s) of the relevant material has/have been appropriately acknowledged after it has been first approved by the courses instructor.
3. In preparing and compiling all this report material, I have not collaborated with anyone and I have not received any type of help from anyone but from the courses instructor.

Signature Michael Person Date 5/03/17