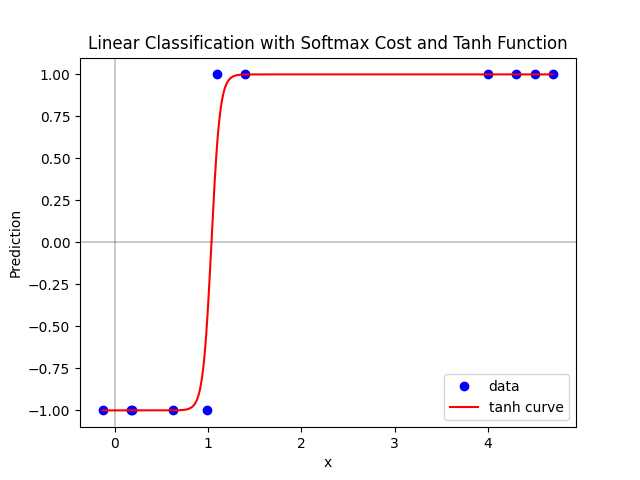
**COMPSCI 589: MACHINE LEARNING**

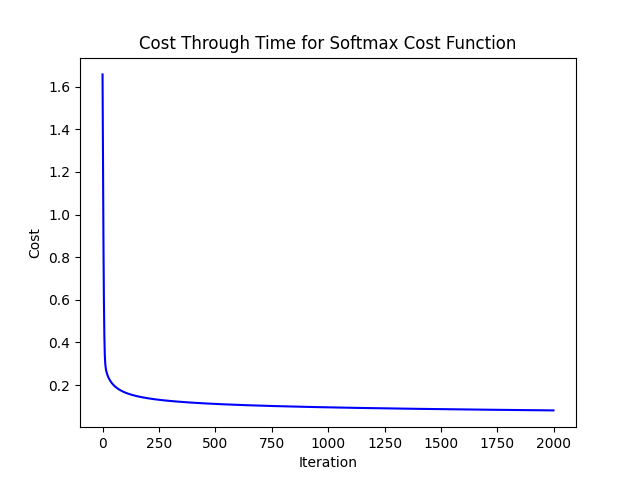
**PROJECT 4 REPORT**

**Student:** Trung Dang – 33858723

1. Task 1

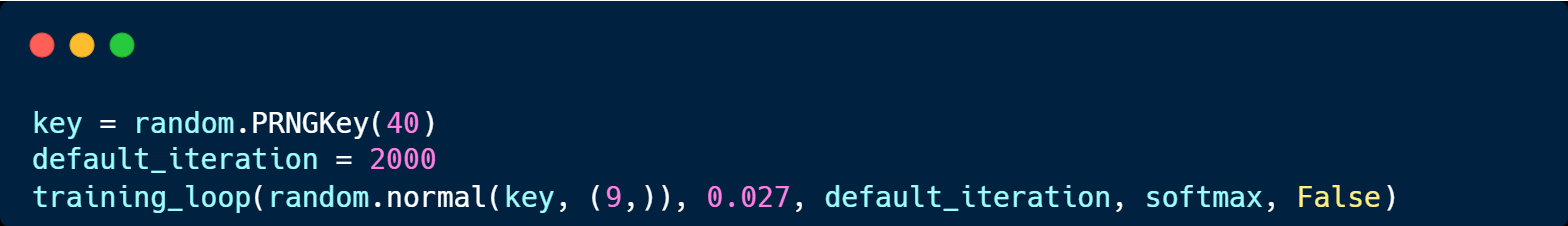


* Accuracy: 100 %
* Misclassification of model = 0
* Figure shows cost history over 2000 iterations.

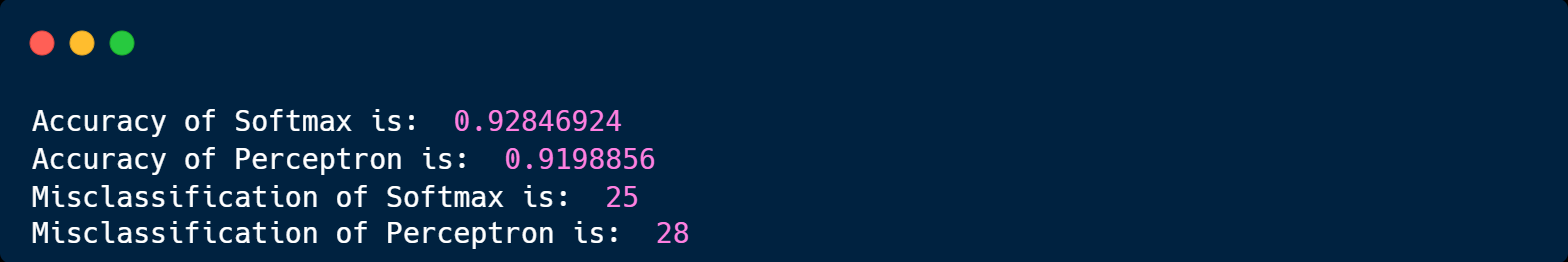


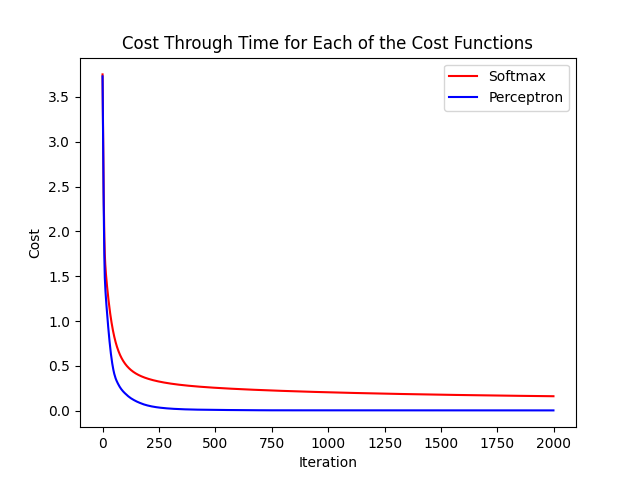
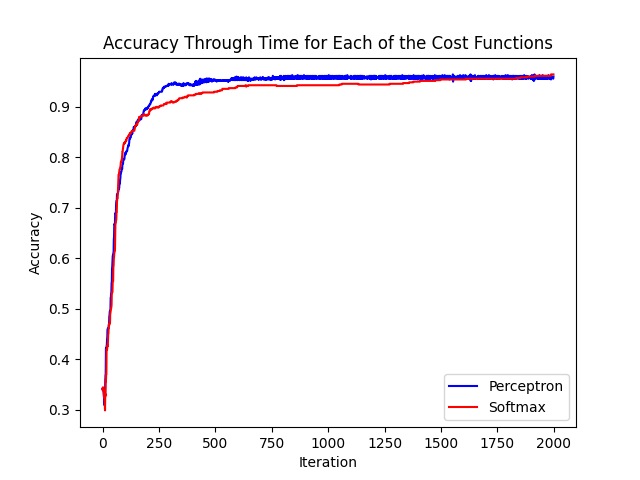
1. Task 2

* Initial values are described by:



* Here, the weights **(w)** are generated as a random array of shape (9,), (8 given parameters + 1 dimension for non-feature-touching weight). To ensure replication between different runs, the seed for the random number generator is set to 40. 40 is a personal preference. In fact, it can be replaced by any constant c.
* The learning rate is chosen to be 0.027. I found that the Perceptron model tends to diverge with large step length. In fact, a step length as small as 0.07 have already caused divergence in the Perceptron model. In [**appendix A**](#Appendix)**,** I have included the graph of the model when run with alpha = 0.07, as well as a table of misclassification of each model when run with different learning rate to demonstrate why 0.027 is approximately optimal
* number of epochs **(max\_it)** = 2000. I have run the same program with 50, 300, 2000, and 5000 epochs. 2000 seems to be the most suitable max\_it, as it allows the model to be refined while do not run into the same aforementioned divergence problem with Perceptron.
* I have explained why this happens in Campuswire post #193 at: <https://campuswire.com/c/G460FD1BA/feed/193>
* `diminishing\_steplength` in the given example is False
* The accuracy and misclassifications from the two cost functions are reported as below:





**APPENDIX A: An** **experiment with learning rates**

|  |  |  |
| --- | --- | --- |
| **Alpha** | **Misclassification** | |
| **Softmax** | **Perceptron** |
| **0.1** | 40 | 33 |
| **0.2** | 33 | 28 |
| **0.25** | 30 | 27 |
| **0.27** | 25 | 28 |
| **0.3** | 25 | 32 |
| **0.4** | 26 | 34 |

Table 1. Misclassification for several alpha

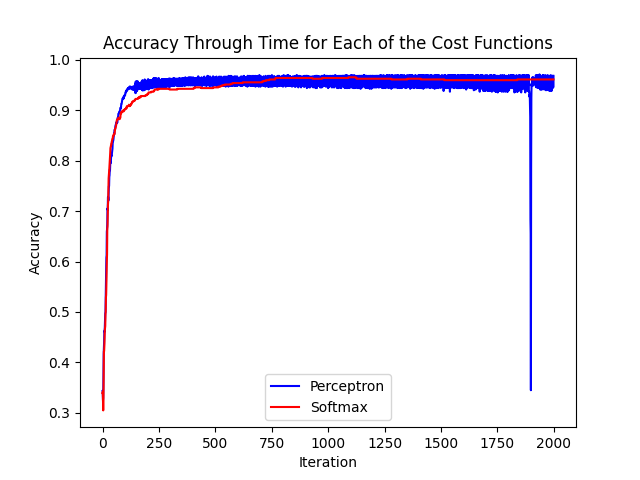
****

Fig 1. Accuracy through time for alpha = 0.07