**David R Mohler**

**EE 5410: Neural Networks**

**Assignment #2**

**02-22-2018**

1. Activation function: Heaviside / Unipolar Discrete
2. Description:
   * In the case of two unique discrete classes which are defined as belonging to either class 1 or class 0 (as per given data) the Heaviside activation function aims for classification in to only these classes with absolute certainty with the goal of minimizing the error in classifications to zero (if classes are linearly separable) .
3. **See cost plots on following page.** 
   * We can observe from the five trials that the choice of the Heaviside activation function causes the jagged fluctuations in cost for the network. For each iteration, as the weights are modified we would hope to see a trend towards a cost of zero. However, due to the binary activation of the output we see oscillation for a large number of epochs until a “sweet spot” is found which yields no errors.
   * The number of epochs varies largely due to the initialization of the weights in the system. We see that each random initialization yields a different number of epochs while no other factors within the network itself are modified.

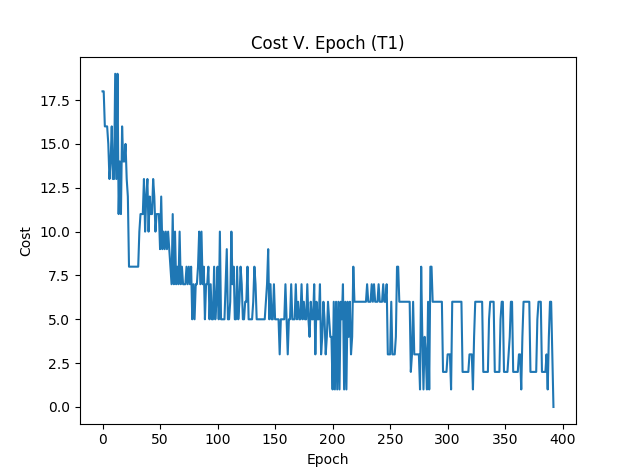
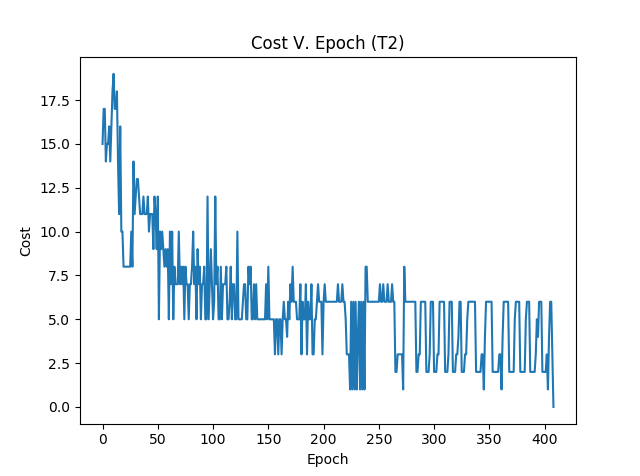
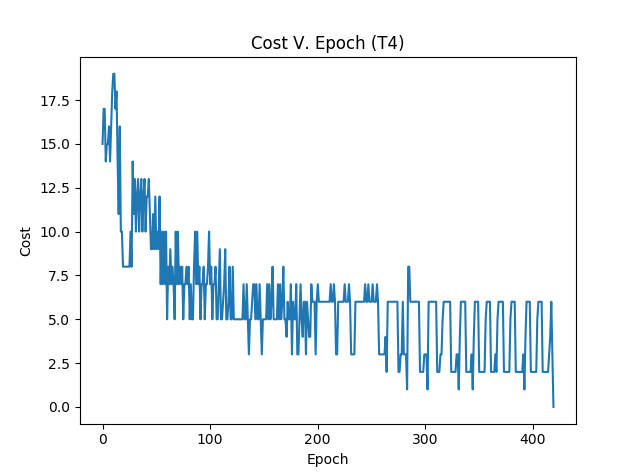
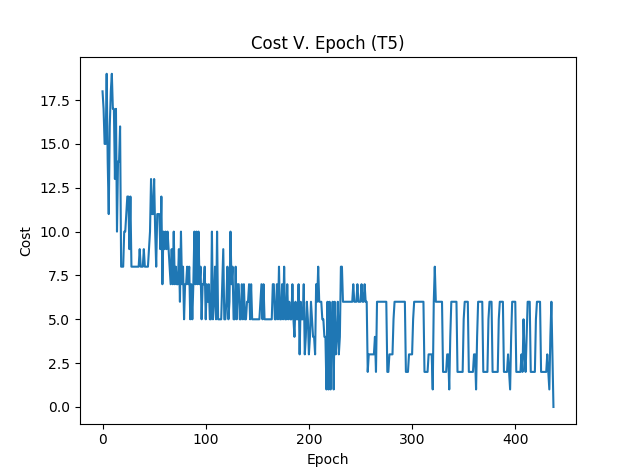
**Table 1 Perceptron Training Results**

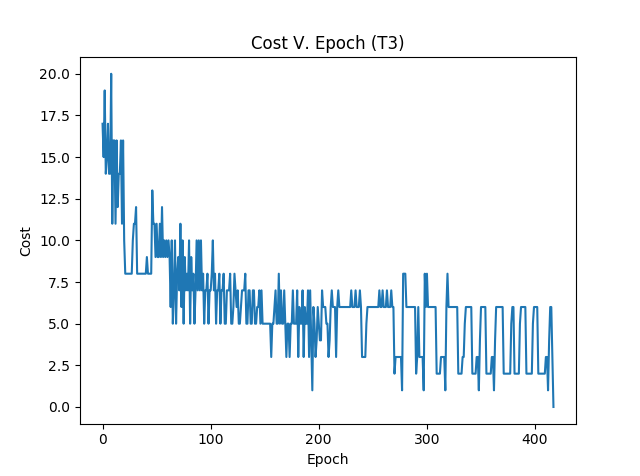
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Training** | **Vector of Weights (Initial)** | | | | **Vector of Weights (Final)** | | | | **# of Epochs** |
| **w1** | **w2** | **w3** | **w0** | **w1** | **w2** | **w3** | **w0** |  |
| **T1** | 0.4412 | 0.4291 | 0.6043 | 0 | 77.46 | 122.51 | -36.26 | -152 | 392 |
| **T2** | 0.6104 | 0.8448 | 0.3861 | 0 | 78.12 | 123.57 | -36.70 | -154 | 408 |
| **T3** | 0.7304 | 0.3064 | 0.8355 | 0 | 78.11 | 124.93 | -37.09 | -156 | 417 |
| **T4** | 0.1228 | 0.5583 | 0.1207 | 0 | 80.96 | 126.33 | -37.29 | -156 | 419 |
| **T5** | 0.3020 | 0.6726 | 0.5010 | 0 | 79.20 | 126.39 | -37.16 | -156 | 437 |

**Table 2: Classification of Test Data**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sample | x1 | x2 | x3 | YP (T1) | YP (T2) | YP (T3) | YP (T4) | YP (T5) |
| 1 | -0.3665 | 0.062 | 5.9891 | 0 | 0 | 0 | 0 | 0 |
| 2 | -0.7842 | 1.1267 | 5.5912 | 1 | 1 | 1 | 1 | 1 |
| 3 | 0.3012 | 0.5611 | 5.8234 | 1 | 1 | 1 | 1 | 1 |
| 4 | 0.7757 | 1.0648 | 8.0677 | 1 | 1 | 1 | 1 | 1 |
| 5 | 0.157 | 0.8028 | 6.304 | 1 | 1 | 1 | 1 | 1 |
| 6 | -0.7014 | 1.0316 | 3.6005 | 1 | 1 | 1 | 1 | 1 |
| 7 | 0.3748 | 0.1536 | 6.1537 | 0 | 0 | 0 | 0 | 0 |
| 8 | -0.692 | 0.9404 | 4.4058 | 1 | 1 | 1 | 1 | 1 |
| 9 | -1.397 | 0.7141 | 4.9263 | 0 | 0 | 0 | 0 | 0 |
| 10 | -1.8842 | -0.2805 | 1.2548 | 0 | 0 | 0 | 0 | 0 |

1. **Equation of the Hypersurface:**
   * The hypersurface separating the two classes is given as the equation of a plane:
   * This equation is derived from the learned weights of the network on any given trial. Using this equation, we are able to view the decision boundary (see classification image on final page). This shows the classifications made by the network on the test patterns simultaneously with the training data, at which point it is obvious the data sets are linearly separable and able to be classified by the single neuron/ perceptron network.





**Legend:**

: Positive Training Data

: Positively Classified Test Data

: Negative Training Data

: Negatively Classified Test Data

