**Implementing. Tuning and Data Curation of Feedforward Networks**

**Bill Yu, CM 2043**

**Basic**

1. 3 training episodes are needed to learn Boolean **and**.
2. threshold is set to 0.2 when only 1 training episode is needed to learn Boolean **and**.
3. learning rate is set to 0.3 or 0.4 when only 1 training episode is needed to learn Boolean **and**.
4. No. It can’t learn when learning rate is 0.3 and the threshold is 0.1.
5. They are contradict to each other.
6. It fails to learn XOR. The reason being first layer functions make sure that at least one element is "TRUE" (using OR) and make sure that they're not all "TRUE" (using NAND). second layer function Make sure that both first-layer criteria are satisfied (using AND). The network's output is just the result of this second function. The first layer transforms the inputs into something that the second layer can use so that the whole network can perform XOR.

**XOR Experiments**



In my opinion, reliable is determined by the error between actual and desired output. when the error is less than 1%, the learning is considered reliable. According to my observation, LearningRate and training episodes are inversely proportional while the increase hiddenLayerSize has no obvious positive effect on the reliability.

|  |  |  |  |
| --- | --- | --- | --- |
| Set | hiddenLayerSize | Learning Rate | episodes |
| 1 | 4 | 0.3 | 100000 |
| 2 | 6 | 3 | 10000 |
| 3 | 6 | 10 | 5000 |
| 4 | 12 | 10 | 5000 |

Set 1:

Example 0 has: 0.007997484871392086 should be: 0

Example 1 has: 0.9917895815225052 should be: 1

Example 2 has: 0.9916582550501286 should be: 1

Example 3 has: 0.008215409927225942 should be: 0

Set 2:

Example 0 has: 0.008068405543634034 should be: 0

Example 1 has: 0.992172520027419 should be: 1

Example 2 has: 0.9913638998900098 should be: 1

Example 3 has: 0.009910188437088865 should be: 0

Set 3:

Example 0 has: 0.005540644248129192 should be: 0

Example 1 has: 0.9952488761373764 should be: 1

Example 2 has: 0.9936655458048337 should be: 1

Example 3 has: 0.006665045667798197 should be: 0

Set4:

Example 0 has: 0.004030868269104064 should be: 0

Example 1 has: 0.9953459842415483 should be: 1

Example 2 has: 0.9945538682823193 should be: 1

Example 3 has: 0.00694383950396862 should be: 0

For Step activation function, the desired output and actual output should be the same when we consider the reliability. I have noticed the training episodes quantity is significantly decreased when using step activation function. However, other relationships is the same to question 10.

|  |  |  |  |
| --- | --- | --- | --- |
| Set | hiddenLayerSize | Learning Rate | episodes |
| 1 | 5 | 0.5 | 1000 |
| 2 | 5 | 0.1 | 1000 |
| 3 | 3 | 0.9 | 1000 |
| 4 | 3 | 0.1 | 5000 |

13.

10 training episodes are needed for precision. Precision is determined by number of correct and overall accuracy.

Episodes: 10

Hidden layer: 2

Hidden layer size: 100

Learning rate: 0.4

The test result shows that:

Execution time is 27.93000 seconds

0 errors

Done testing.

Number correct: 10000 out of: 10000

Overall accuracy: 1.0