**MACHINE LEARNING**

Q1 a.

Overview –

Learning rate - random between 0 to 0.5

Stopping condition - Which ever comes first -

1. If predicted output differs from actual output by less than 0.1 (for all outputs)

2. If we have run 100k training iterations

Two Methods of ANN –

1. In one I’ve learnt based only 3 hidden nodes, where the ANN learnt 7 nodes correctly.

2. In another I’ve learnt based on 3 hidden nodes and a **learning** **BIAS** on each hidden node and outputs. I’ve observed that with a BIAS, the **ANN learnt all 8 nodes.** I believe the reason is because the BIAS acts like an extra hidden node, but one without input to it.

1. Without BIAS – 7 node convergence

Hidden Vals Rounded Hidden Values

0.10 0.09 0.08 0 0 0

1.00 1.00 0.00 1 1 0

0.45 0.08 1.00 0 0 1

0.04 0.47 0.99 0 0 1

1.00 0.27 0.27 1 0 0

0.27 1.00 0.19 0 1 0

0.00 1.00 0.74 0 1 1

1.00 0.00 0.98 1 0 1

Outputs Vals Rounded Outputs

0.39 0.07 0.03 0.09 0.07 0.06 0.04 0.05 0 0 0 0 0 0 0 0

0.05 0.98 0.00 0.00 0.00 0.00 0.00 0.00 0 1 0 0 0 0 0 0

0.06 0.00 0.95 0.05 0.00 0.00 0.00 0.04 0 0 1 0 0 0 0 0

0.06 0.00 0.04 0.94 0.00 0.00 0.04 0.00 0 0 0 1 0 0 0 0

0.09 0.00 0.00 0.00 0.97 0.00 0.00 0.01 0 0 0 0 1 0 0 0

0.08 0.02 0.00 0.00 0.00 0.97 0.02 0.00 0 0 0 0 0 1 0 0

0.04 0.00 0.00 0.04 0.00 0.03 0.96 0.00 0 0 0 0 0 0 1 0

0.04 0.00 0.03 0.00 0.01 0.00 0.00 0.96 0 0 0 0 0 0 0 1

1. With BIAS – 8 Node convergence

Hidden Vals Rounded Hidden Values

0.17 0.00 0.56 0 0 1

0.99 0.05 0.97 1 0 1

0.01 0.78 0.41 0 1 0

0.56 0.99 0.02 1 1 0

0.26 0.24 0.01 0 0 0

0.99 0.36 0.06 1 0 0

0.18 0.46 1.00 0 0 1

0.97 0.99 0.97 1 1 1

Outputs Vals Rounded Outputs

0.91 0.04 0.02 0.00 0.06 0.00 0.07 0.00 1 0 0 0 0 0 0 0

0.04 0.95 0.00 0.00 0.00 0.04 0.02 0.04 0 1 0 0 0 0 0 0

0.00 0.00 0.92 0.06 0.02 0.00 0.05 0.01 0 0 1 0 0 0 0 0

0.00 0.00 0.04 0.93 0.04 0.04 0.00 0.05 0 0 0 1 0 0 0 0

0.07 0.00 0.05 0.03 0.90 0.04 0.00 0.00 0 0 0 0 1 0 0 0

0.00 0.04 0.00 0.03 0.05 0.94 0.00 0.01 0 0 0 0 0 1 0 0

0.04 0.01 0.06 0.00 0.00 0.00 0.91 0.04 0 0 0 0 0 0 1 0

0.00 0.02 0.00 0.00 0.00 0.00 0.03 0.93 0 0 0 0 0 0 0 1

Q1b.

As an encoder decoder, the ANN can be used to compress data. In the above example, we have 8 features, and would typically require 8 bits to represent the data. But using a encoder, decoder scheme, we could represent the same 8 bits as 3 bits (Hidden Nodes).

The input-hidden section would be the encoder section, where we encode 8 bits into 3.

The hidden-output section would be the decoder section, where we decode 3 bits into 8.

This sort of encoding-decoding strategy is useful for transferring information over the Internet (compression) and forms the basis of information theory.

Q1c.

No this encoder-decoder scheme will not work for with 1 or 2 hidden nodes for the above 8 inputs.

In information theory, it has been proved that the best compression (encoding-decoding) scheme cannot compress beyond –

Entropy = lg n [where lg is log to the base 2]

We have 8 inputs/outputs (or 8 possible states), so,

Entropy = lg(8) = 3

Thus we cannot correctly encode-decode (compress & uncompress) the inputs/outputs with less than 3 hidden nodes.

Q2a. & Q2c Tables and AUC values for corresponding classifiers-

Each row in the tables below corresponds to 1 iteration in K-folds where K = 10

|  |  |  |
| --- | --- | --- |
| **Bern -** |  |  |
| FP | FN | Err |
| 11 | 30 | 8.89 |
| 16 | 34 | 10.87 |
| 15 | 46 | 13.26 |
| 11 | 39 | 10.87 |
| 7 | 32 | 8.48 |
| 11 | 38 | 10.65 |
| 12 | 32 | 9.57 |
| 12 | 26 | 8.26 |
| 13 | 30 | 9.35 |
| 16 | 38 | 11.74 |
| **avg error** | **10.19** |  |
| **avg AUC** | **0.96151** |  |
|  |  |  |
| **Gauss -** |  |  |
| FP | FN | Err |
| 71 | 5 | 16.49 |
| 80 | 5 | 18.48 |
| 71 | 11 | 17.83 |
| 70 | 8 | 16.96 |
| 83 | 4 | 18.91 |
| 68 | 9 | 16.74 |
| 70 | 9 | 17.17 |
| 76 | 7 | 18.04 |
| 72 | 8 | 17.39 |
| 74 | 11 | 18.48 |
| **avg error** | **17.65** |  |
| **avg AUC** | **0.950911** |  |
|  |  |  |
| **Histogram** |  |  |
| FP | FN | Err |
| 7 | 30 | 8.03 |
| 18 | 36 | 11.74 |
| 15 | 42 | 12.39 |
| 11 | 34 | 9.78 |
| 7 | 31 | 8.26 |
| 16 | 38 | 11.74 |
| 9 | 37 | 10 |
| 11 | 26 | 8.04 |
| 12 | 33 | 9.78 |
| 20 | 32 | 11.3 |
| **avg error** | **10.11** |  |
| **avg AUC** | **0.96727** |  |

Q2b. ROC Curve

