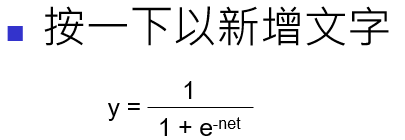
***Q3(ai) Rewrite the above table such that values “Yes” and “No” in attribute “Buy\_NintendoLabo” are mapped to values 1 and 0, respectively.***

|  |  |  |
| --- | --- | --- |
| No. of Computers | No. of Phones | Buy\_NintendoLabo |
| 2 | 0 | 0 |
| 0 | 2 | 0 |
| 4 | 2 | 1 |
| 2 | 4 | 1 |

***Q3(aii)***

α = 0.6



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| b | w1 | w2 | net | y | bnew | w1new | w2new |
| 0.300000 | 0.300000 | 0.300000 | 0.900000 | 0.710950 | -0.126570 | -0.553139 | 0.300000 |
| -0.126570 | -0.553139 | 0.300000 | 0.473430 | 0.616195 | -0.496287 | -0.553139 | -0.439434 |
| -0.496287 | -0.553139 | -0.439434 | -3.587713 | 0.026917 | 0.087563 | 1.782260 | 0.728265 |
| 0.087563 | 1.782260 | 0.728265 | 6.565144 | 0.998593 | 0.088407 | 1.783948 | 0.731641 |

***Q3 (b) What is the major disadvantage of the traditional neural network model compared with the recurrent neural network model?***

Standard RNNs (Recurrent Neural Networks) suffer from vanishing and exploding gradient problems. LSTMs (Long Short Term Memory) deal with these problems by introducing new gates, such as input and forget gates, which allow for a better control over the gradient flow and enable better preservation of “long-range dependencies”.

***Q5 (a) Consider the traditional LSTM model. (i) What are the values of the above status variables when t = 1 and when t = 2? Please show each***

***answer up to 4 decimal places.***

1. t = 1
   1. forget gate variable ft

* 1. input gate variable it
  2. input activation variable at
  3. internal state variable st
  4. output gate variable ot
  5. final output variable yt

1. t = 2
   1. forget gate variable ft
   2. input gate variable it
   3. input activation variable at
   4. internal state variable st
   5. output gate variable ot
   6. final output variable yt

***Q5 (a)*** ***(ii) What are the values of the above status variables when t = 1 and when t = 2? Please show each***

***answer up to 4 decimal places.***

When t=1, Error = y1 – y =

When t=2, Error = y2 – y =

***Q5 (b) Consider the GRU model (i) What are the values of the above status variables when t = 1 and when t = 2? Please show each***

***answer up to 4 decimal places.***

1. t = 1
   1. reset gate variable rt
   2. input activation variable at
   3. update gate variable ut
   4. final output variable yt
2. t = 2
   1. reset gate variable rt
   2. input activation variable at
   3. update gate variable ut
   4. final output variable yt

***Q5 (b)*** ***(ii) What are the values of the above status variables when t = 1 and when t = 2? Please show each***

***answer up to 4 decimal places.***

When t=1, Error = y1 – y =

When t=2, Error = y2 – y =

***Q5 (b)*** ***(c) What are the advantages and the disadvantages of the GRU model compared with the traditional LSTM model?***

it can be difficult to train standard RNNs to solve problems that require learning long-term temporal dependencies. This is because the gradient of the loss function decays exponentially with time (called the vanishing gradient problem). LSTM networks are a type of RNN that uses special units in addition to standard units. LSTM units include a 'memory cell' that can maintain information in memory for long periods of time. A set of gates is used to control when information enters the memory, when it's output, and when it's forgotten. This architecture lets them learn longer-term dependencies. GRUs are similar to LSTMs, but use a simplified structure. They also use a set of gates to control the flow of information, but they don't use separate memory cells, and they use fewer gates.