**Exercise #2**

In Exercise 1, the single-layer feed-forward neural network was trained for 495 epochs, achieving a final training error of 0.04477200082565811. When tested with the input [0.1, 0.2], it produced an output of 0.408845.

While in Exercise 2, a two-layer feed-forward network was trained using a gradient descent backpropagation for 1000 epochs, with a final error of 0.09679992618680057. Testing the same input gave an output of 0.50169851.

The results show that the more complex network architecture in Exercise 2 were more effective in learning the underlying pattern in the data compared to the simpler single-layer network architecture in Exercise 1.

**Exercise #3**

**Epoch Errors:**

* The error values during training are much higher compared to Exercise 1, with an initial error around **2.85**.
* The error gradually reduces but fluctuates throughout training, showing signs of instability.
* The final error value is **1.51**, which is significantly higher than the final error in Exercise 1.

**Test Result for [0.1, 0.2] with 100 training instances:**

* The predicted value is **0.45347296**, closer to the expected value **0.3** compared to Exercise 1 **(0.681)** but still not highly accurate.

**Key Observations:**

* The network architecture with 6 neurons in hidden layer may not have been sufficient to effectively learn the larger dataset as we saw in Exercise 3.
* Exercise 1 with only 10 training instances, the model likely overfit the small dataset, leading to poor performance on new input.
* Exercise 3 with 100 training instances, the model had more data to learn from and provided answer closer to (0.3).
* Exercise 3 shows the benefit of training with a larger dataset.

**Exercise #5**

This exercise asks for the code to implement a feed-forward neural network to recognize a sum pattern with 3 inputs. It goes through generating training data, training a neural network, testing then re-training with added data to give an analysis improvement in performance. With limited data the network may have issues recognizing a pattern in an accurate manner which may cause deviations in the predicted data, but increasing the size of the dataset it allows the network to perform better with improvement in accuracy of result 6 compared to result 5.