Using Multi-layer perceptron denoising signal

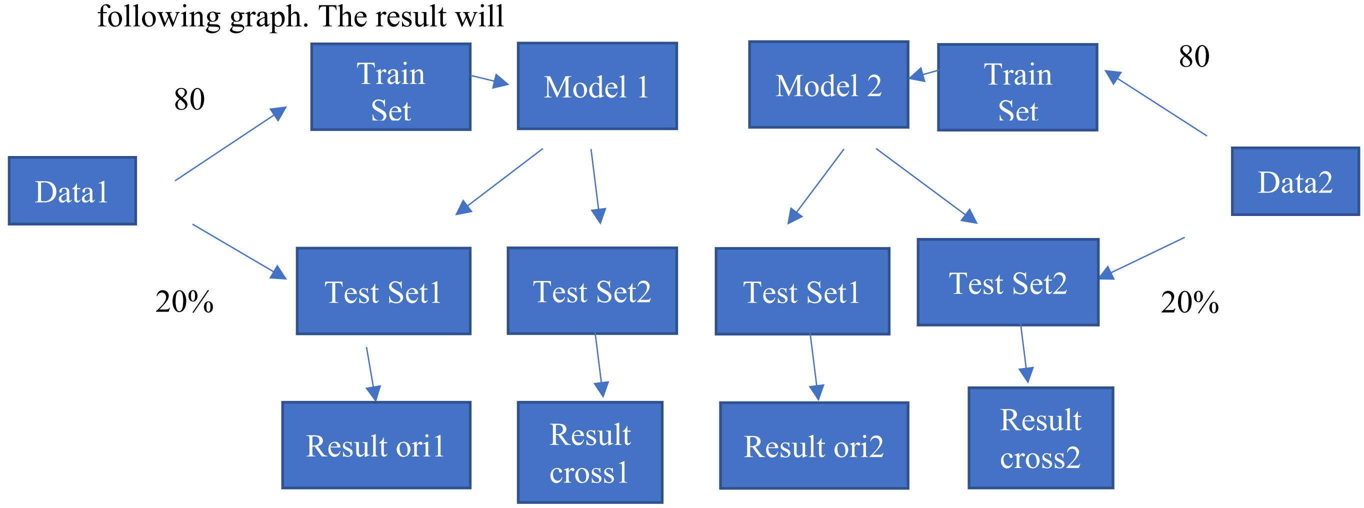
**Introduction**

This report discusses an experiment of used Multi-layer perceptron as a learning model to predict the true signal symbol. The experiment is to test the following property:

- If you increase the size of the test data set, has this error probability reached a "steady state"?
- If you repeat with different simulated training sets of the same size, how does the error distribution change when comparing to the same test data set(s)?
- If you increase the size of the simulated training set, how does this impact the error distribution?

**Procedures**

- Experiment I used 10 data sets, with same data distribution but different sample size. The first data set has 50 samples in each block and totally has 1,000 blocks (50*1,000 samples), and the following data set has 1,000 more blocks comparing with the previous data set. Therefore, the last data set has 50 * 10,000 samples. The test result will compare the accuracy in each data set, and the error distribution.
- For testing property 2, Experiment II used 2 different data set and each data set has 50*1000 samples, named them data 1 and data 2 respectively, and the procedures as the following graph. The result will



will compare the accuracy and the error distribution from data1 and data2 by using Euclidean distance to compute the closest neighbor.

**Results and discussion**

When increasing the testing size, the error rate has reached a steady state. As the following table shows, the accuracy rate has converged, although the data set size has increased 10 times.

| Increased size test | |
|---|---|
| Data set | Accuracy |
| 1 | 0.727 |
| 2 | 0.731 |
| 3 | 0.733 |
| 4 | 0.716 |
| 5 | 0.732 |
| 6 | 0.733 |
| 7 | 0.735 |
| 8 | 0.722 |
| 9 | 0.732 |
| 10 | 0.739 |

When using the training model to test a sample with different distribution, the error distribution will differ with its original result. On the other hand, the error distribution result is more similar with the different distribution error result. The mean Euclidean distance between the result_ori1 and result_cross2 is 0.27, and the mean Euclidean distance between the result_ori2 and result_cross1 is 0.039.

When increased the training set, the error distribution didn't affect by the sample size. However, the error distribution also converged to a "steady state."

**Conclusion**

In this experiment, I used 2 main distribution, which has large scale noise or interference, so the model didn't show a high performance on predicting result. The reason is I want to leverage the difference in the error distribution, so the result can show clearly. As I observe the Euclidean Distance table, there is a large difference between result_ori1 with other result, and the reason caused this result should because the mean error of result_ori1(19% loss) is lower than other result (30% loss or more).

**Appendix**

Experiment I:

- Sample distribution: "SNR: 60, INR: 5, P_INT: 0.8, N*30."
- Accuracy distribution

| Data set | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 1 | 0.931 | 0.955 | 0.957 | 0.925 | 0.543 | 0.534 |
| 2 | 0.956 | 0.960 | 0.951 | 0.943 | 0.652 | 0.515 |
| 3 | 0.950 | 0.957 | 0.959 | 0.938 | 0.695 | 0.521 |
| 4 | 0.956 | 0.946 | 0.942 | 0.953 | 0.651 | 0.522 |
| 5 | 0.954 | 0.959 | 0.946 | 0.942 | 0.671 | 0.497 |
| 6 | 0.942 | 0.958 | 0.959 | 0.949 | 0.677 | 0.518 |
| 7 | 0.952 | 0.944 | 0.945 | 0.959 | 0.656 | 0.508 |
| 8 | 0.960 | 0.943 | 0.946 | 0.959 | 0.648 | 0.483 |
| 9 | 0.950 | 0.947 | 0.952 | 0.952 | 0.692 | 0.530 |
| 10 | 0.948 | 0.961 | 0.957 | 0.945 | 0.651 | 0.472 |

Experiment I:

- Sample distribution:

- Data1: (“SNR*30, INR*30, P_INT*0.8, N*1)
- Data2: (“SNR*60, INR*5, P_INT*0.8, N*30)

- Euclidean distance

| | Result_cross1 | Result_cross2 |
|---|---|---|
| Result_ori1 | 0.34 | 0.27 |
| Result_ori2 | 0.039 | 0.05 |