Artificial Neural Network

Homework 1
(Multi-Layer Perceptron)

Course Instructor:
Dr. Rezaeian

Amir Mohammad Taghizadehgan

Student Number: 40104244

Summer 2023

Q.1) data set:

```
(506, 14)
```

	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	0.00632	18.0	2.31	0	0.538	6.575	65.2	4.0900	1	296.0	15.3	396.90	4.98	24.0
1	0.02731	0.0	7.07	0	0.469	6.421	78.9	4.9671	2	242.0	17.8	396.90	9.14	21.6
2	0.02729	0.0	7.07	0	0.469	7.185	61.1	4.9671	2	242.0	17.8	392.83	4.03	34.7
3	0.03237	0.0	2.18	0	0.458	6.998	45.8	6.0622	3	222.0	18.7	394.63	2.94	33.4
4	0.06905	0.0	2.18	0	0.458	7.147	54.2	6.0622	3	222.0	18.7	396.90	5.33	36.2

for normalization, we simply subtract each column from its own average, and then divide the result by its std. As we have split data to train and test, we have to normalize test and train separately.

Since we want to use 5-fold cross validation method, we set number of epochs to 5. And the batch size is set to 1 for this problem. The results for each fold are acquired as shown:

```
MAE = [2.4024754, 2.6389174, 2.7718284, 3.3717327, 3.195421]
(Mean Absolute Error)
```

And the result for test is:

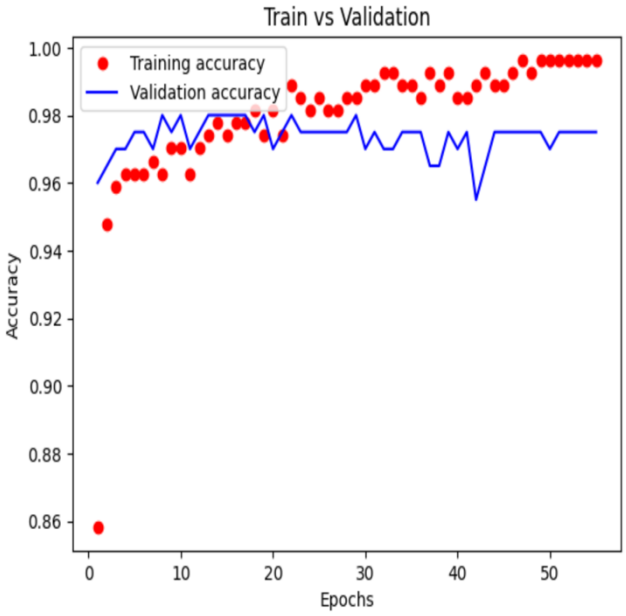
```
MSE = 28.15542221069336, MAE: 3.932856321334839
(Mean Squared Error)
```

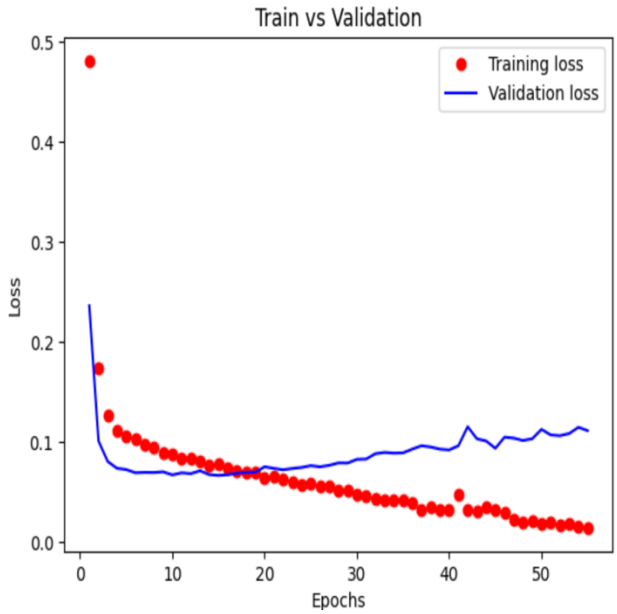
Q.2) data set:

```
(699, 10)
```

	1	2	3	4	5	6	7	8	9	10
0	5	1	1	1	2	1	3	1	1	2
1	5	4	4	5	7	10	3	2	1	2
2	3	1	1	1	2	2	3	1	1	2
3	6	8	8	1	3	4	3	7	1	2
4	4	1	1	3	2	1	3	1	1	2

Normalization is just like Q.1. and here are the model evaluation plots:



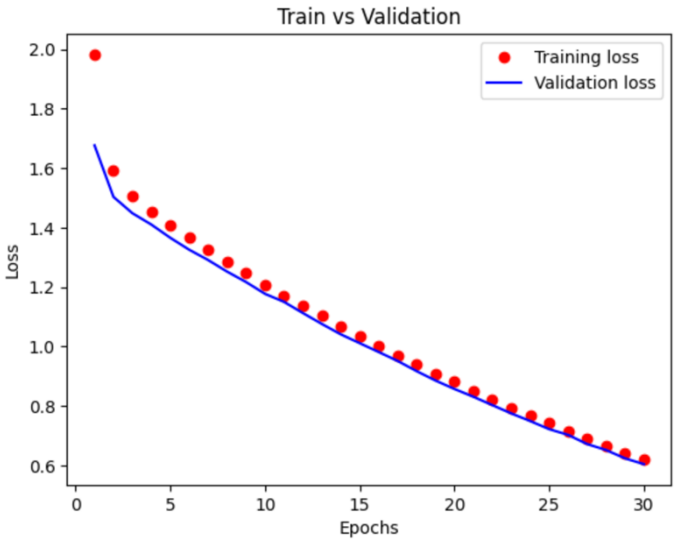


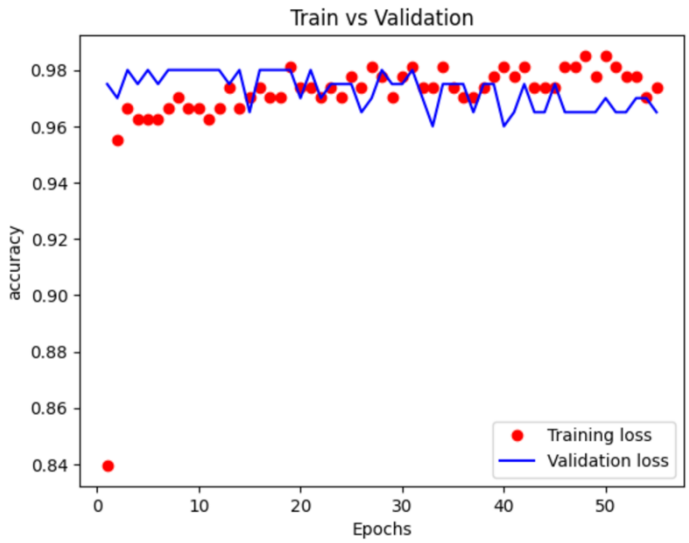
4 epochs is the optimal number of epochs.

With optimal number of epochs, we train our model again and here are the test evaluation results:

```
MSE = 0.11008329689502716, MAE: 0.9567099809646606
```

But when we add L1 regularization with λ = 0.001 to all layers, with 55 epochs we get:






```
MSE = 0.3222477436065674, MAE: 0.939393937587738
```

Q.3)

ReLu: for inputs below zero, this activation function may lead to vanishing gradient. Because for $z<0$, a will be equal to zero.

Tanh: Because the slope of tanh in high and low values are almost equal to zero, for small and large values, vanishing gradient happens.

Leaky ReLu: this activation function is one of the solutions when vanishing gradient happens. Because it doesn't suffer from $z<0$ like ReLu.