

# CS 4375: Neural Net

# ASSIGNMENT 2

Names of students in your group:

JGR190002 Joseph Ramos, DYS190002 David Song

Number of free late days used: 0

Note: You are allowed a total of 4 free late days for the entire semester. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Please list clearly all the sources/references that you have used in this assignment.

## Part 1

### Assignment 2

1.1

$$O = w_0 + w_1(x_1 + x_1^2) + \dots + w_n(x_n + x_n^2)$$

$$\text{Error} = E = \frac{1}{2}(O - t)^2$$

$$\frac{\partial E}{\partial w_i} = \frac{\partial E}{\partial O} \times \frac{\partial O}{\partial w_i} \quad w_i \text{ is the weight}$$

$$\frac{\partial E}{\partial O} = E \quad , \quad \frac{\partial O}{\partial w_i} = x_i + x_i^2$$

Using the chain rule,

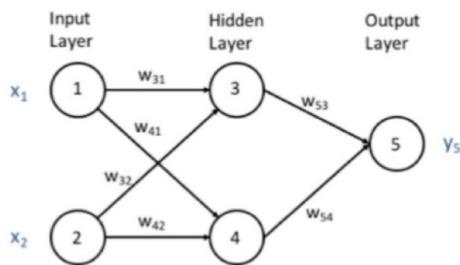
$$\Delta w_i = -\eta \left( \frac{\partial E}{\partial O} \right)$$

$$\left( \frac{\partial E}{\partial O} \right) \Delta w_i = -\eta \times E \times (x_i + x_i^2)$$

$$\therefore w_{i,\text{new}} = w_{i,\text{old}} - \eta E (x_i + x_i^2)$$

6-2

Q.



for the first neuron of hidden layer which is indicated as 3.

$$z_1 = \text{Neuron 3} = (x_1 \times w_{31}) + (x_2 \times w_{32})$$
$$h_1 = h(z_1)$$

for the second neuron of hidden layer which is indicated as 4.

$$z_2 = \text{Neuron 4} = (x_1 \times w_{41}) + (x_2 \times w_{42})$$
$$h_2 = h(z_2)$$

The output of the neural network which is  $y_5$ .

$$y_5 = (w_{53} \times h_1) + (w_{54} \times h_2)$$

$$= h(w_{53} \cdot h(w_{31} \cdot x_1 + w_{32} \cdot x_2)) + \\ (w_{54} \cdot h(w_{41} \cdot x_1 + w_{42} \cdot x_2))$$

b.

$$\text{Input vector } X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$W^{(1)} = \begin{pmatrix} w_{31} & w_{32} \\ w_{41} & w_{42} \end{pmatrix}$$

$$W^{(2)} = [w_{53}, w_{54}]$$

$$\text{Input of hidden layer} = W^{(1)} \cdot X$$

$$= \begin{bmatrix} w_{31} & w_{32} \\ w_{41} & w_{42} \end{bmatrix}_{2 \times 2} \times \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_{2 \times 1} = \begin{bmatrix} w_{31}x_1 + w_{32}x_2 \\ w_{41}x_1 + w_{42}x_2 \end{bmatrix}$$

$$Z = \text{output of hidden layer} = h(W^{(1)} \cdot X)$$

$$\text{input of output layer} = W^{(2)} \cdot h(W^{(1)} \cdot X)$$

$$= \begin{bmatrix} w_{53} & w_{54} \end{bmatrix}_{1 \times 2} \cdot \begin{bmatrix} w_{31}x_1 + w_{32}x_2 \\ w_{41}x_1 + w_{42}x_2 \end{bmatrix}_{2 \times 1}$$

$$= \begin{bmatrix} w_{53}(w_{31}x_1 + w_{32}x_2) + w_{54}(w_{41}x_1 + w_{42}x_2) \end{bmatrix}_{1 \times 1}$$

$$Y = \text{output of output layer} = h(W^{(2)} \cdot h(W^{(1)} \cdot X))$$

$$= h(w_{53}(w_{31}x_1 + w_{32}x_2) + w_{54}(w_{41}x_1 + w_{42}x_2))$$

$$\text{C. Sigmoid} = S = \frac{1}{1+e^{-x}} \\ = \frac{e^x}{1+e^x}$$

$$S + S \cdot e^x = 1$$

$$e^x = \frac{1-S}{S}$$

$$\text{Tanh} = T = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{\left(\frac{1-S}{S}\right) - \left(\frac{S}{1-S}\right)}{\left(\frac{1-S}{S}\right) + \left(\frac{S}{1-S}\right)}$$

Sub  
S in

$$= \frac{1-2S}{1-2S+2S^2}$$

Simplif.

$$= -\frac{(1-e^{-x})(1+e^{-x})}{(1+e^{-2x})}$$

Sub  
back  
in

## Trial history of every combination:

### **VERY BIG NOTE:**

With the way we coded our for loop, the program iterates through every single possible combination of hyper-parameters and shows you the tabular results after. In every iteration it displays a graph. Iterations go from 100-400, and 100-900. This does not mean that iteration 500 or 1000 is not done, it just is not displayed. The program does take a bit to run, so be patient please.

For our hyperparameters, we used our own set of values with the same parameters given from the .py by the professor. Since it provides more adequate outputs to compare between the combinations, and overall shows better performance.

We set the values to compare as

```
activations = ['logistic', 'tanh', 'relu']
learning_rates = [0.1, 0.01]
max_iters = [500, 1000]
hidden_layer_size = [(10, 15), (20, 35)]
```

Data Split: 80% training 20% testing

I believe the assignment specified to give the link for the dataset:

- Link to the dataset used. *Please do not include the data as part of your submission.*

Dataset Link:

<https://utdallas.box.com/shared/static/b7woqah5sk2ooh7od6xw0sd163p0kajm.csv>

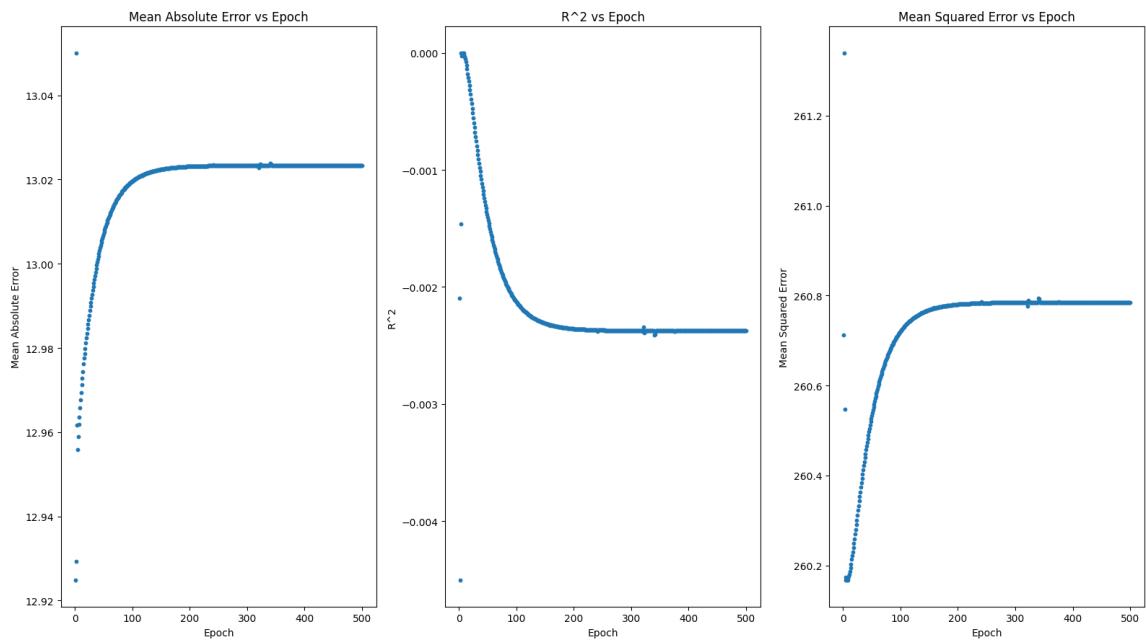
We are starting with the fixed hidden layer size of (10,15) and changing other parameters for trial 0-11. Hidden layer size of (20,35) and different combinations for trials 12-23.

Combination 0:

```

Hidden layer size: (10, 15) Activation Func: logistic Learning Rate: 0.1 Epoch/Iterations: 500
Epoch 100: MAE = 13.0197, R2 = -0.0021, MSE = 260.7227
Epoch 200: MAE = 13.0231, R2 = -0.0024, MSE = 260.7807
Epoch 300: MAE = 13.0233, R2 = -0.0024, MSE = 260.7846
Epoch 400: MAE = 13.0233, R2 = -0.0024, MSE = 260.7848

```

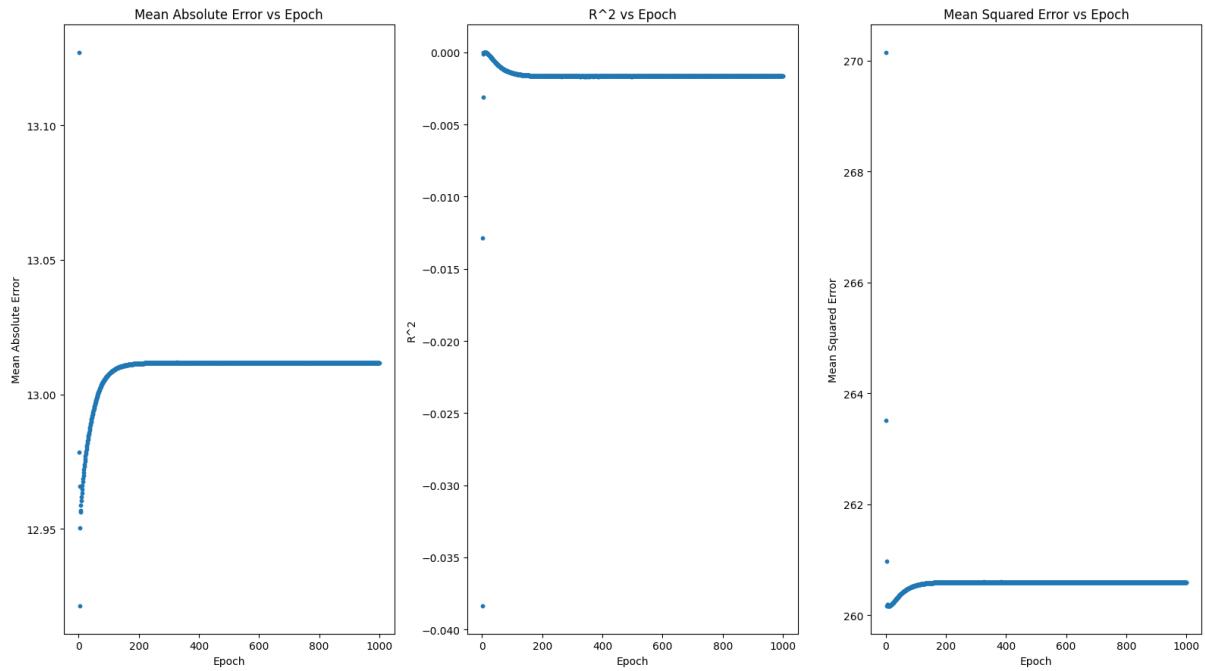


### combination 1:

```

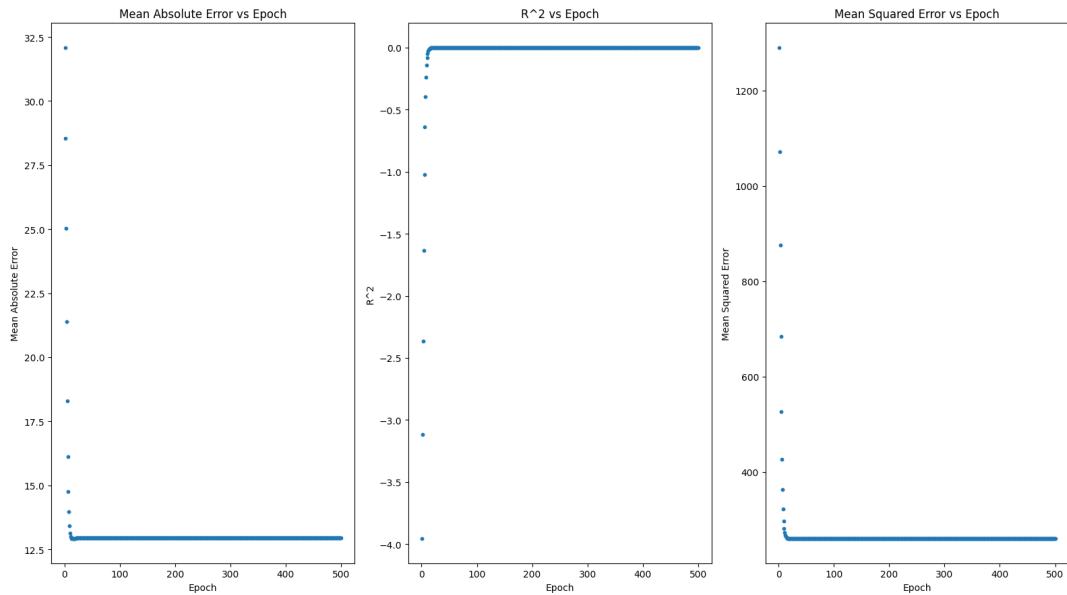
Hidden layer size: (10, 15) Activation Func: logistic Learning Rate: 0.1 Epoch/Iterations: 1000
Epoch 100: MAE = 13.0076, R2 = -0.0014, MSE = 260.5402
Epoch 200: MAE = 13.0115, R2 = -0.0016, MSE = 260.5944
Epoch 300: MAE = 13.0117, R2 = -0.0017, MSE = 260.5982
Epoch 400: MAE = 13.0117, R2 = -0.0017, MSE = 260.5985
Epoch 500: MAE = 13.0117, R2 = -0.0017, MSE = 260.5984
Epoch 600: MAE = 13.0117, R2 = -0.0017, MSE = 260.5984
Epoch 700: MAE = 13.0117, R2 = -0.0017, MSE = 260.5983
Epoch 800: MAE = 13.0117, R2 = -0.0017, MSE = 260.5983
Epoch 900: MAE = 13.0117, R2 = -0.0017, MSE = 260.5983

```



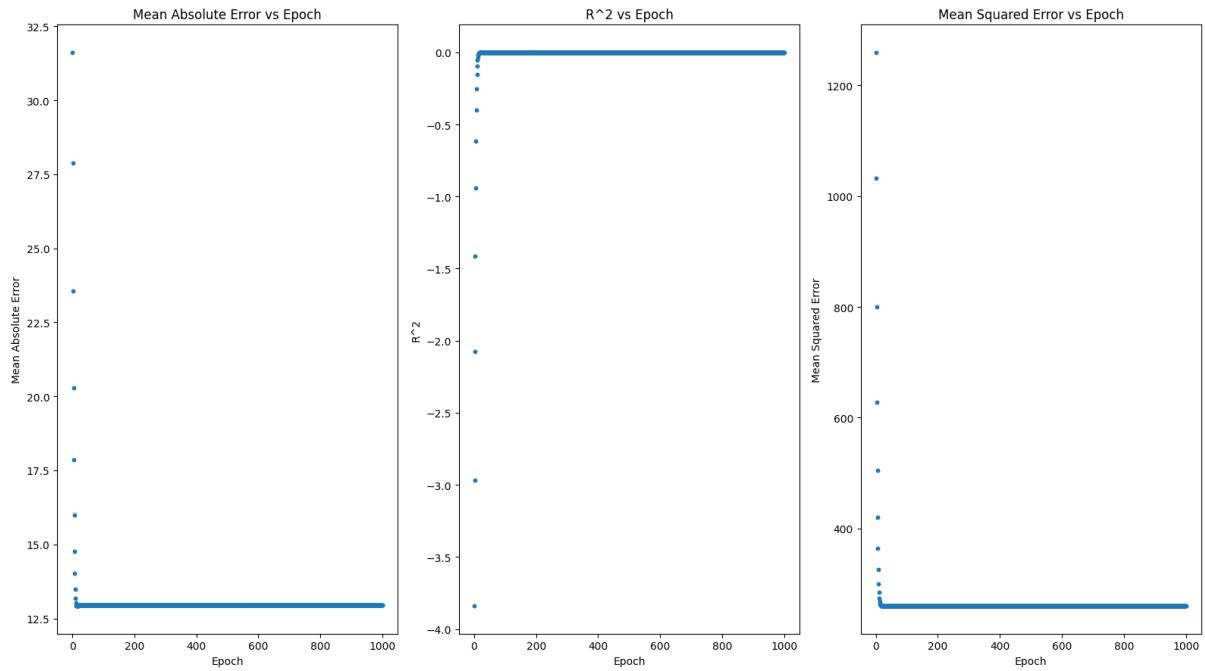
### Combination 2:

```
Hidden layer size: (10, 15) Activation Func: logistic Learning Rate: 0.01 Epoch/Iterations: 500
Epoch 100: MAE = 12.9414, R2 = -0.0005, MSE = 260.2985
Epoch 200: MAE = 12.9420, R2 = -0.0005, MSE = 260.2901
Epoch 300: MAE = 12.9421, R2 = -0.0005, MSE = 260.2878
Epoch 400: MAE = 12.9420, R2 = -0.0005, MSE = 260.2896
```



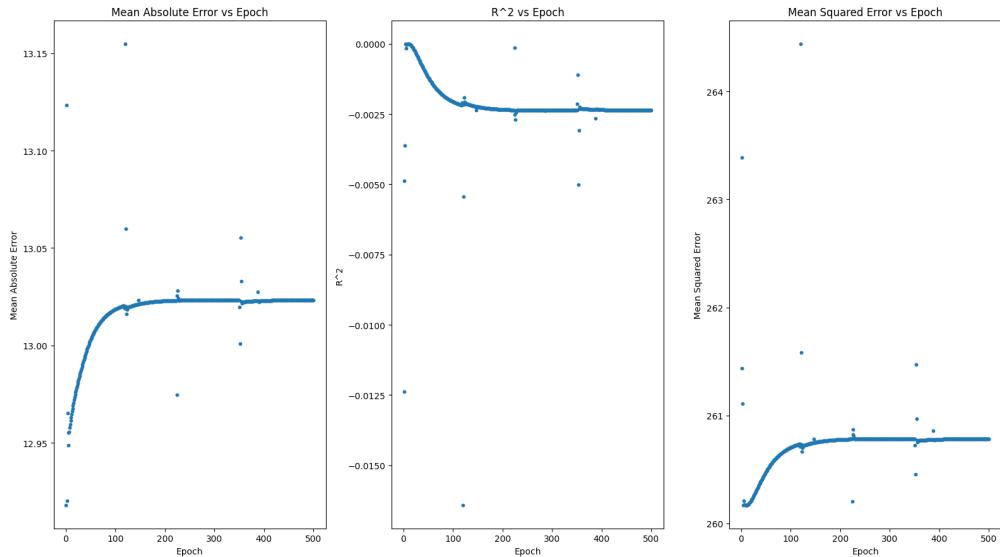
### combination 3:

```
Hidden layer size: (10, 15) Activation Func: logistic Learning Rate: 0.01 Epoch/Iterations: 1000
Epoch 100: MAE = 12.9414, R2 = -0.0005, MSE = 260.2993
Epoch 200: MAE = 12.9419, R2 = -0.0005, MSE = 260.2910
Epoch 300: MAE = 12.9425, R2 = -0.0004, MSE = 260.2824
Epoch 400: MAE = 12.9420, R2 = -0.0005, MSE = 260.2901
Epoch 500: MAE = 12.9420, R2 = -0.0005, MSE = 260.2899
Epoch 600: MAE = 12.9420, R2 = -0.0005, MSE = 260.2896
Epoch 700: MAE = 12.9421, R2 = -0.0005, MSE = 260.2889
Epoch 800: MAE = 12.9420, R2 = -0.0005, MSE = 260.2897
Epoch 900: MAE = 12.9421, R2 = -0.0005, MSE = 260.2889
```



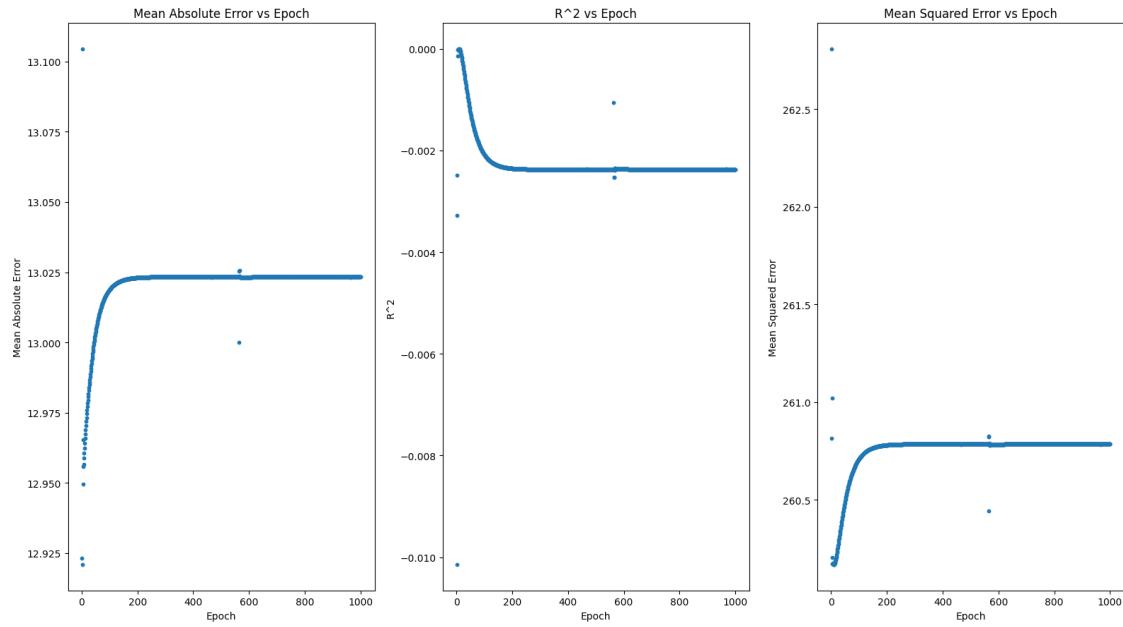
#### combination 4:

```
Hidden layer size: (10, 15) Activation Func: tanh Learning Rate: 0.1 Epoch/Iterations: 500
Epoch 100: MAE = 13.0187, R2 = -0.0021, MSE = 260.7058
Epoch 200: MAE = 13.0228, R2 = -0.0023, MSE = 260.7759
Epoch 300: MAE = 13.0233, R2 = -0.0024, MSE = 260.7841
Epoch 400: MAE = 13.0230, R2 = -0.0023, MSE = 260.7781
```



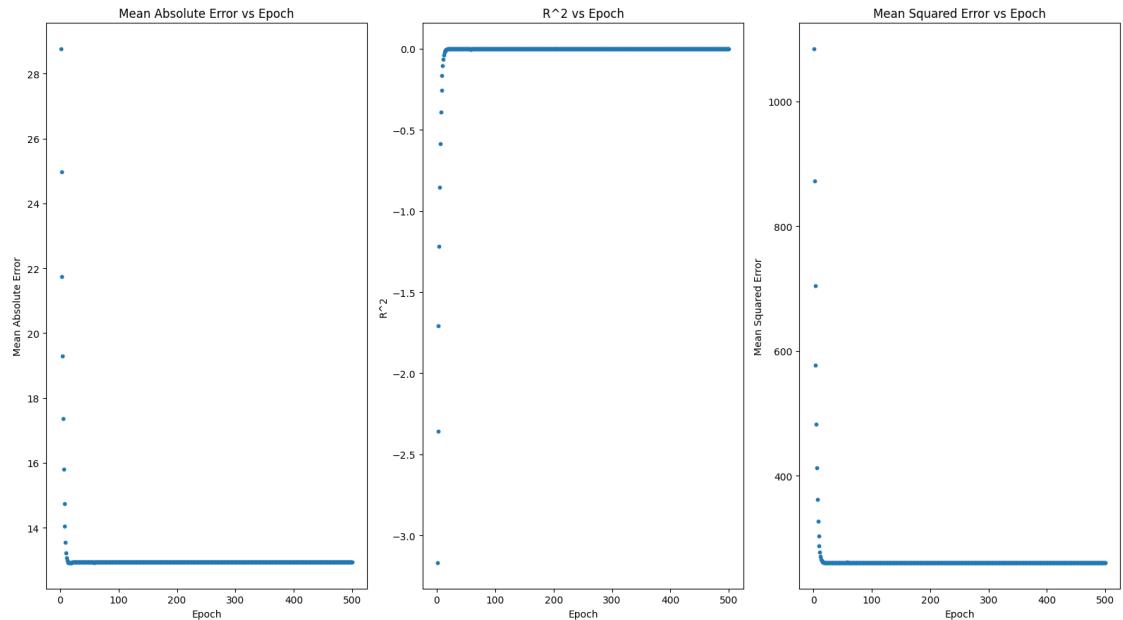
#### combination 5:

```
Hidden layer size: (10, 15) Activation Func: tanh Learning Rate: 0.1 Epoch/Iterations: 1000
Epoch 100: MAE = 13.0190, R2 = -0.0021, MSE = 260.7107
Epoch 200: MAE = 13.0231, R2 = -0.0024, MSE = 260.7797
Epoch 300: MAE = 13.0233, R2 = -0.0024, MSE = 260.7842
Epoch 400: MAE = 13.0233, R2 = -0.0024, MSE = 260.7845
Epoch 500: MAE = 13.0233, R2 = -0.0024, MSE = 260.7845
Epoch 600: MAE = 13.0232, R2 = -0.0024, MSE = 260.7826
Epoch 700: MAE = 13.0233, R2 = -0.0024, MSE = 260.7844
Epoch 800: MAE = 13.0233, R2 = -0.0024, MSE = 260.7845
Epoch 900: MAE = 13.0233, R2 = -0.0024, MSE = 260.7845
```



combination 6:

```
Hidden layer size: (10, 15) Activation Func: tanh Learning Rate: 0.01 Epoch/Iterations: 500
Epoch 100: MAE = 12.9417, R2 = -0.0005, MSE = 260.2948
Epoch 200: MAE = 12.9419, R2 = -0.0005, MSE = 260.2910
Epoch 300: MAE = 12.9420, R2 = -0.0005, MSE = 260.2900
Epoch 400: MAE = 12.9420, R2 = -0.0005, MSE = 260.2901
```

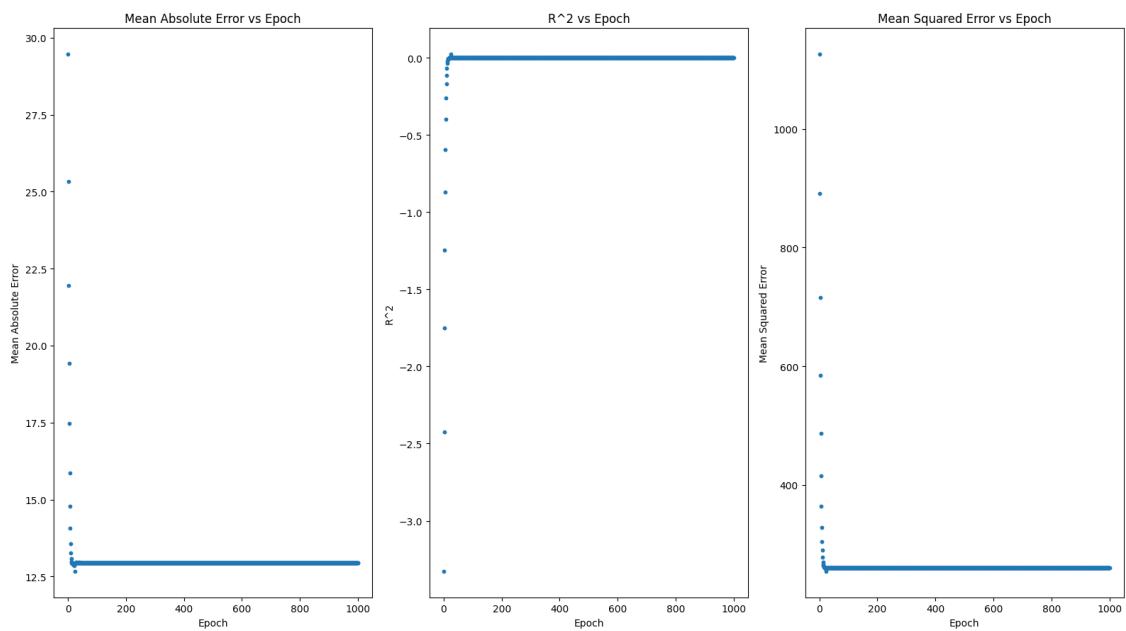


combination 7:

```

Hidden layer size: (10, 15) Activation Func: tanh Learning Rate: 0.01 Epoch/Iterations: 1000
Epoch 100: MAE = 12.9429, R2 = -0.0005, MSE = 260.2954
Epoch 200: MAE = 12.9426, R2 = -0.0004, MSE = 260.2817
Epoch 300: MAE = 12.9425, R2 = -0.0004, MSE = 260.2820
Epoch 400: MAE = 12.9425, R2 = -0.0004, MSE = 260.2822
Epoch 500: MAE = 12.9425, R2 = -0.0004, MSE = 260.2824
Epoch 600: MAE = 12.9425, R2 = -0.0004, MSE = 260.2823
Epoch 700: MAE = 12.9420, R2 = -0.0005, MSE = 260.2900
Epoch 800: MAE = 12.9420, R2 = -0.0005, MSE = 260.2901
Epoch 900: MAE = 12.9420, R2 = -0.0005, MSE = 260.2901

```

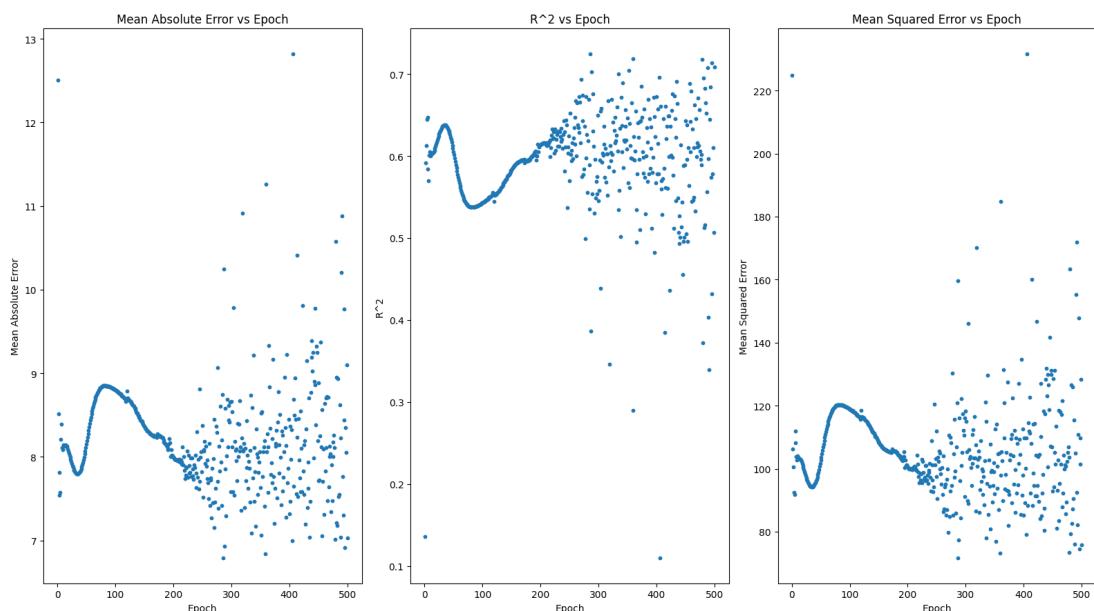


combination 8:

```

Hidden layer size: (10, 15) Activation Func: relu Learning Rate: 0.1 Epoch/Iterations: 500
Epoch 100: MAE = 8.7938, R2 = 0.5435, MSE = 118.7581
Epoch 200: MAE = 7.9849, R2 = 0.6141, MSE = 100.4016
Epoch 300: MAE = 8.5386, R2 = 0.5582, MSE = 114.9335
Epoch 400: MAE = 8.3659, R2 = 0.5783, MSE = 109.7014

```

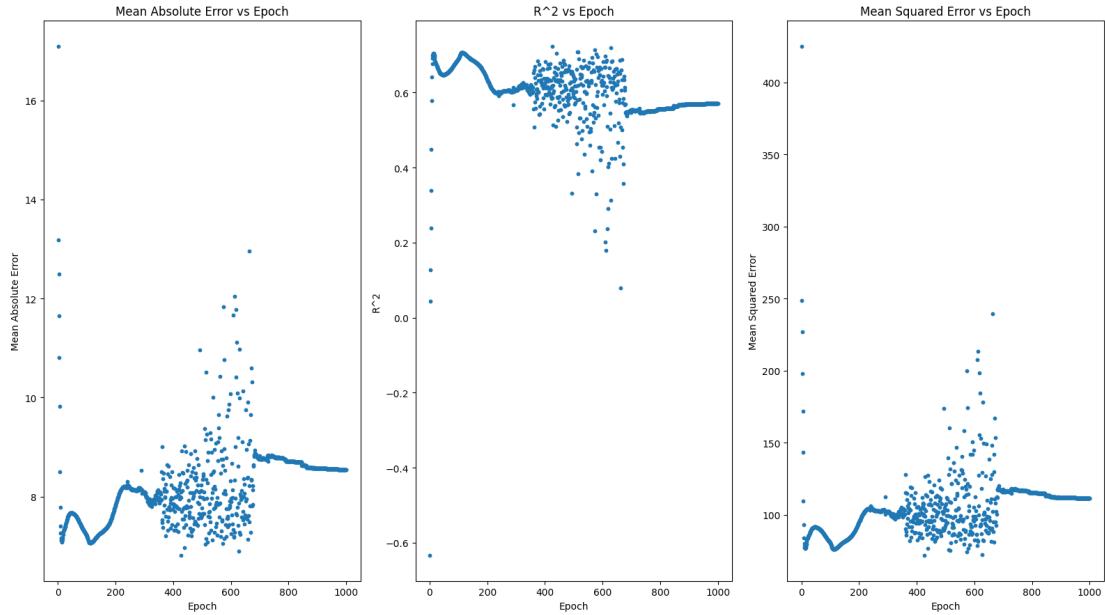


combination 9:

```

Hidden layer size: (10, 15) Activation Func: relu Learning Rate: 0.1 Epoch/Iterations: 1000
Epoch 100: MAE = 7.2029, R2 = 0.6921, MSE = 80.1089
Epoch 200: MAE = 7.8489, R2 = 0.6332, MSE = 95.4392
Epoch 300: MAE = 8.0900, R2 = 0.6059, MSE = 102.5356
Epoch 400: MAE = 8.1801, R2 = 0.5948, MSE = 105.4286
Epoch 500: MAE = 7.6530, R2 = 0.6391, MSE = 93.8932
Epoch 600: MAE = 7.2486, R2 = 0.6825, MSE = 82.5915
Epoch 700: MAE = 8.7854, R2 = 0.5512, MSE = 116.7660
Epoch 800: MAE = 8.7190, R2 = 0.5567, MSE = 115.3439
Epoch 900: MAE = 8.5743, R2 = 0.5690, MSE = 112.1266

```

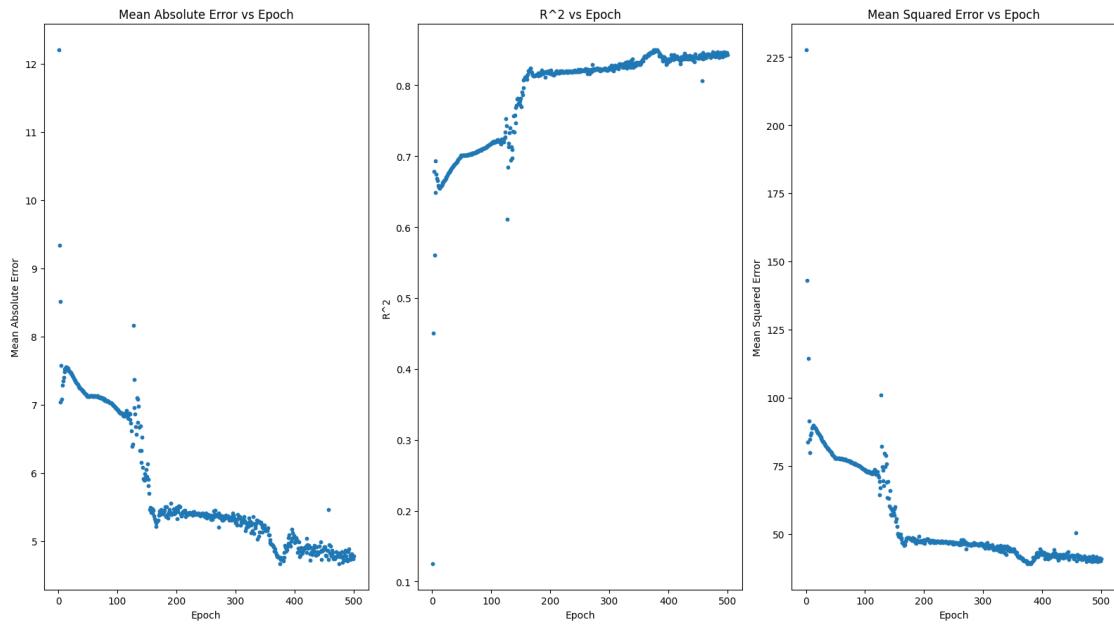


combination 10:

```

Hidden layer size: (10, 15) Activation Func: relu Learning Rate: 0.01 Epoch/Iterations: 500
Epoch 100: MAE = 6.9305, R2 = 0.7177, MSE = 73.4513
Epoch 200: MAE = 5.3312, R2 = 0.8210, MSE = 46.5699
Epoch 300: MAE = 5.2611, R2 = 0.8256, MSE = 45.3861
Epoch 400: MAE = 4.9885, R2 = 0.8368, MSE = 42.4549

```

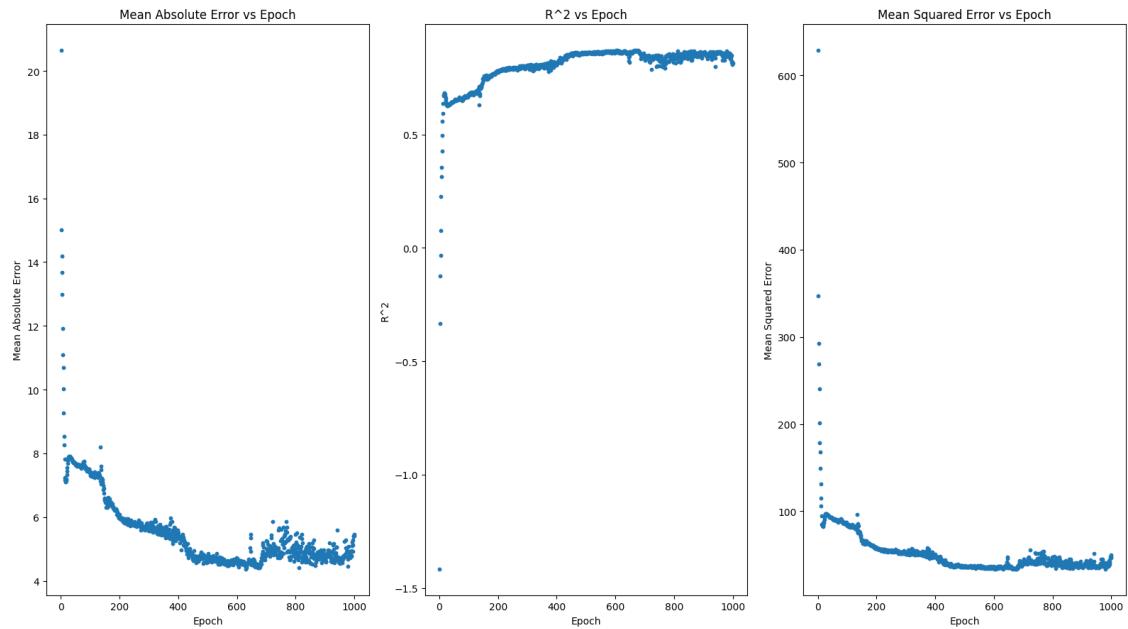


combination 11:

```

Hidden layer size: (10, 15) Activation Func: relu Learning Rate: 0.01 Epoch/Iterations: 1000
Epoch 100: MAE = 7.3035, R2 = 0.6810, MSE = 83.0022
Epoch 200: MAE = 5.9662, R2 = 0.7806, MSE = 57.0726
Epoch 300: MAE = 5.6016, R2 = 0.8022, MSE = 51.4586
Epoch 400: MAE = 5.5024, R2 = 0.8106, MSE = 49.2875
Epoch 500: MAE = 4.7232, R2 = 0.8595, MSE = 36.5625
Epoch 600: MAE = 4.5869, R2 = 0.8662, MSE = 34.8096
Epoch 700: MAE = 5.2565, R2 = 0.8263, MSE = 45.1933
Epoch 800: MAE = 5.2192, R2 = 0.8327, MSE = 43.5285
Epoch 900: MAE = 4.6411, R2 = 0.8623, MSE = 35.8318

```

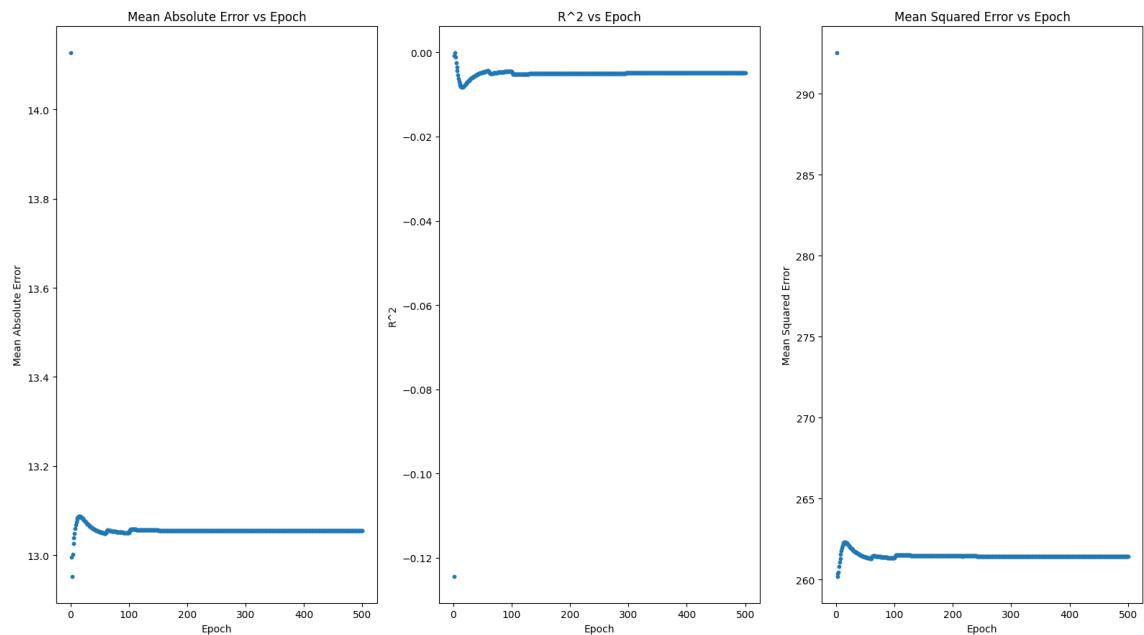


combination 12(Now with fixed hidden layer size of (20,35) and changing other parameters):

```

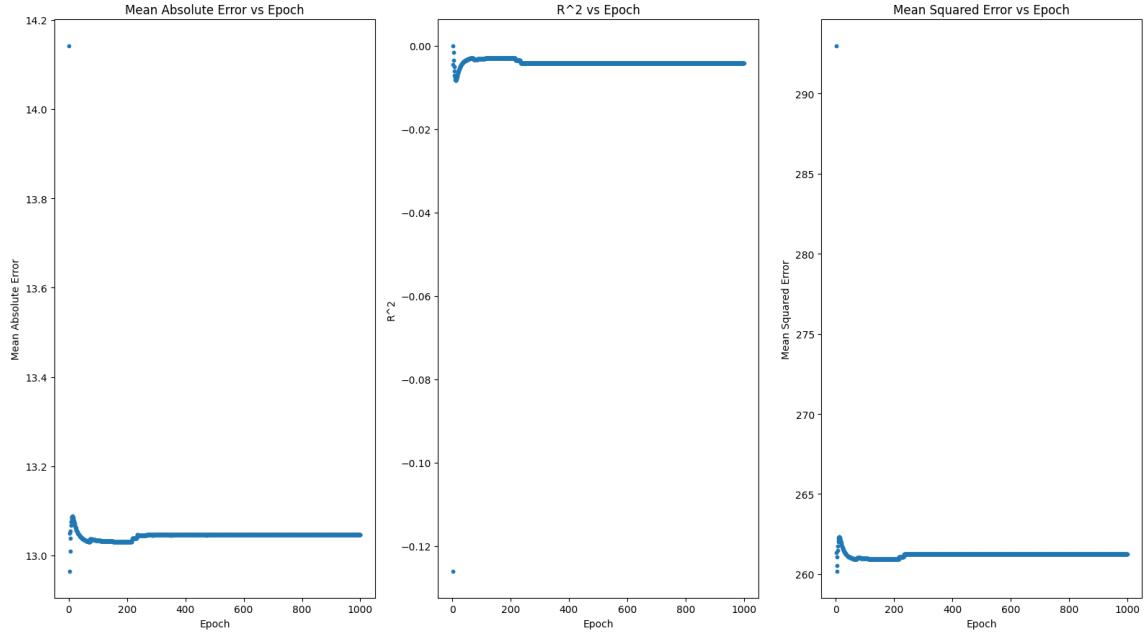
Hidden layer size: (20, 35) Activation Func: logistic Learning Rate: 0.1 Epoch/Iterations: 500
Epoch 100: MAE = 13.0534, R2 = -0.0048, MSE = 261.4216
Epoch 200: MAE = 13.0550, R2 = -0.0050, MSE = 261.4625
Epoch 300: MAE = 13.0547, R2 = -0.0049, MSE = 261.4547
Epoch 400: MAE = 13.0547, R2 = -0.0049, MSE = 261.4542

```



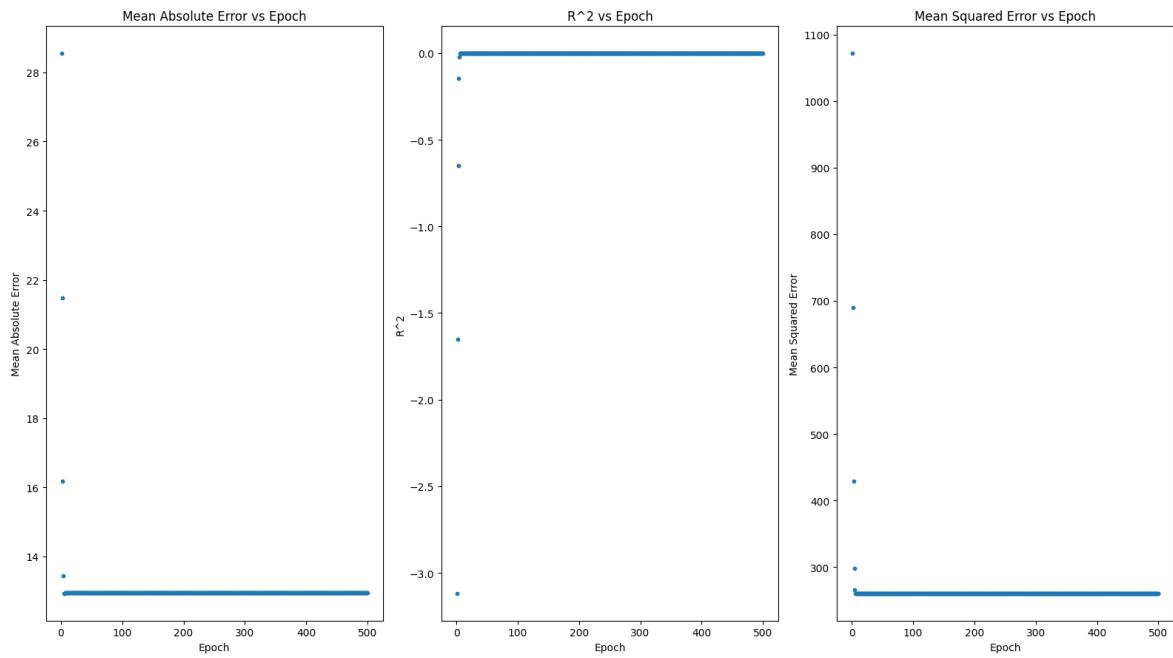
### combination 13:

```
Hidden layer size: (20, 35) Activation Func: logistic Learning Rate: 0.1 Epoch/Iterations: 1000
Epoch 100: MAE = 13.0335, R2 = -0.0031, MSE = 260.9743
Epoch 200: MAE = 13.0307, R2 = -0.0029, MSE = 260.9198
Epoch 300: MAE = 13.0458, R2 = -0.0041, MSE = 261.2391
Epoch 400: MAE = 13.0458, R2 = -0.0041, MSE = 261.2394
Epoch 500: MAE = 13.0458, R2 = -0.0041, MSE = 261.2394
Epoch 600: MAE = 13.0458, R2 = -0.0041, MSE = 261.2395
Epoch 700: MAE = 13.0458, R2 = -0.0041, MSE = 261.2396
Epoch 800: MAE = 13.0458, R2 = -0.0041, MSE = 261.2397
Epoch 900: MAE = 13.0458, R2 = -0.0041, MSE = 261.2397
```



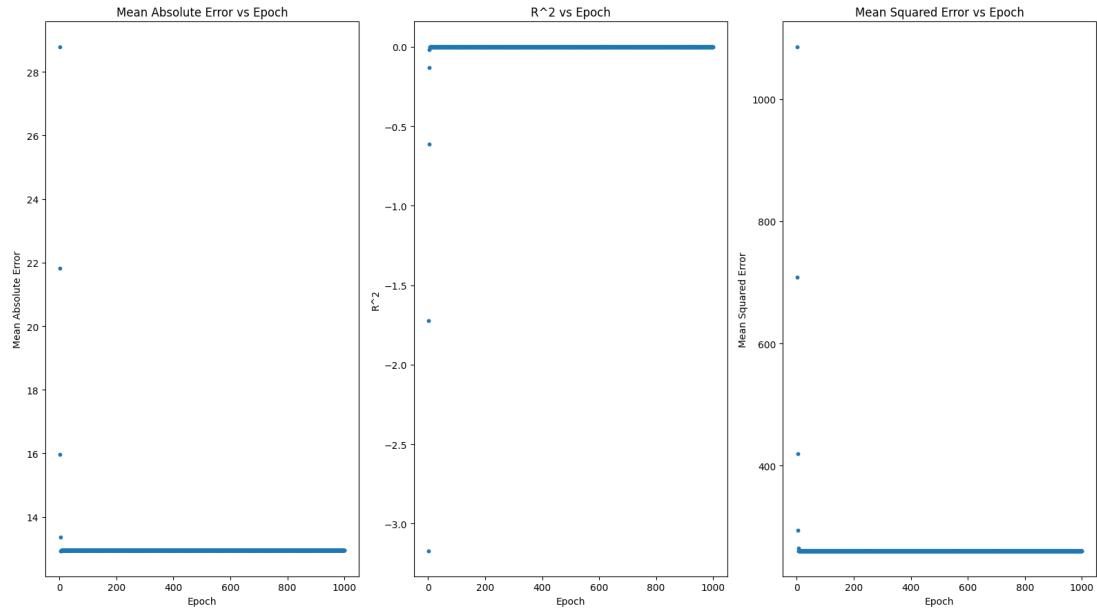
### combination 14:

```
Hidden layer size: (20, 35) Activation Func: logistic Learning Rate: 0.01 Epoch/Iterations: 500
Epoch 100: MAE = 12.9454, R2 = -0.0003, MSE = 260.2453
Epoch 200: MAE = 12.9459, R2 = -0.0003, MSE = 260.2395
Epoch 300: MAE = 12.9459, R2 = -0.0003, MSE = 260.2394
Epoch 400: MAE = 12.9459, R2 = -0.0003, MSE = 260.2396
```



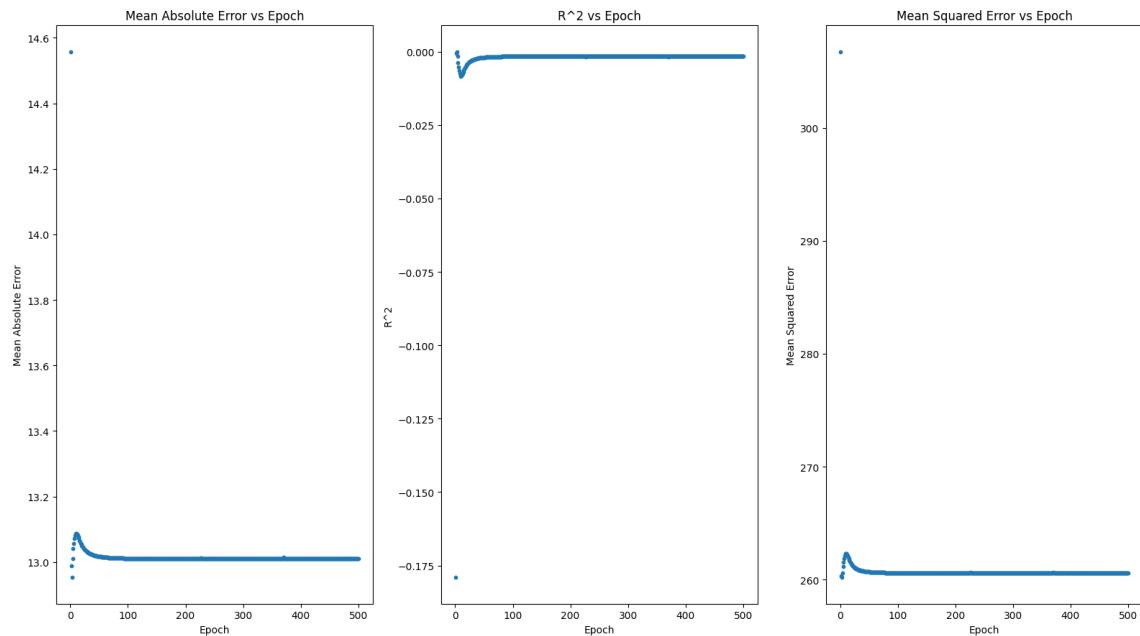
### combination 15:

```
Hidden layer size: (20, 35) Activation Func: logistic Learning Rate: 0.01 Epoch/Iterations: 1000
Epoch 100: MAE = 12.9451, R2 = -0.0003, MSE = 260.2483
Epoch 200: MAE = 12.9458, R2 = -0.0003, MSE = 260.2401
Epoch 300: MAE = 12.9458, R2 = -0.0003, MSE = 260.2398
Epoch 400: MAE = 12.9458, R2 = -0.0003, MSE = 260.2398
Epoch 500: MAE = 12.9458, R2 = -0.0003, MSE = 260.2398
Epoch 600: MAE = 12.9458, R2 = -0.0003, MSE = 260.2399
Epoch 700: MAE = 12.9458, R2 = -0.0003, MSE = 260.2399
Epoch 800: MAE = 12.9458, R2 = -0.0003, MSE = 260.2398
Epoch 900: MAE = 12.9459, R2 = -0.0003, MSE = 260.2396
```



### combination 16:

```
Hidden layer size: (20, 35) Activation Func: tanh Learning Rate: 0.1 Epoch/Iterations: 500
Epoch 100: MAE = 13.0111, R2 = -0.0016, MSE = 260.5888
Epoch 200: MAE = 13.0100, R2 = -0.0016, MSE = 260.5732
Epoch 300: MAE = 13.0099, R2 = -0.0016, MSE = 260.5722
Epoch 400: MAE = 13.0099, R2 = -0.0016, MSE = 260.5722
```

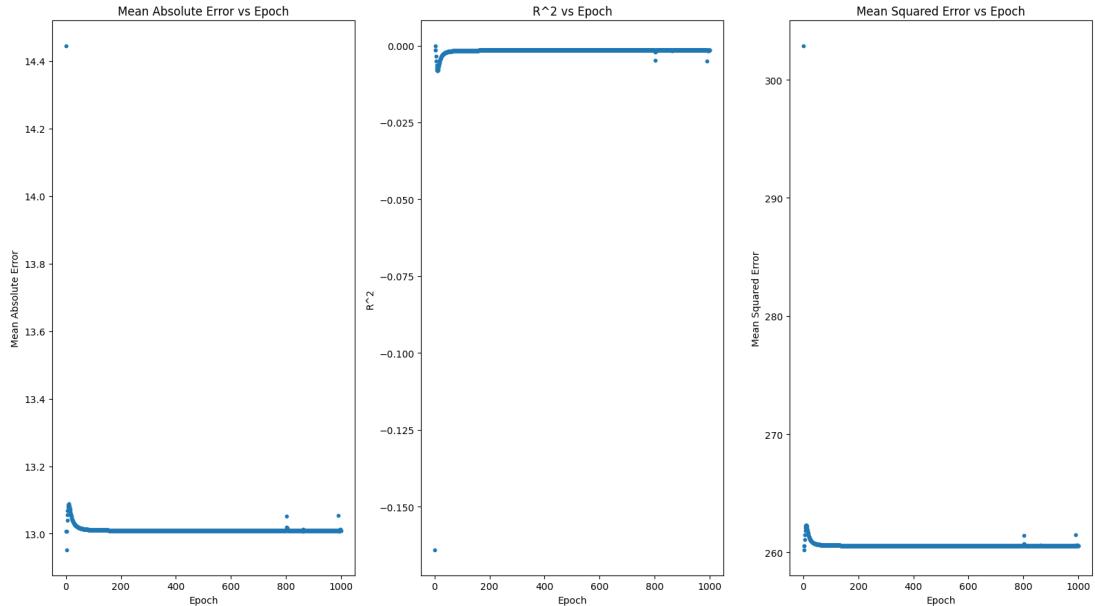


### combination 17:

```

Hidden layer size: (20, 35) Activation Func: tanh Learning Rate: 0.1 Epoch/Iterations: 1000
Epoch 100: MAE = 13.0111, R2 = -0.0016, MSE = 260.5891
Epoch 200: MAE = 13.0100, R2 = -0.0016, MSE = 260.5732
Epoch 300: MAE = 13.0099, R2 = -0.0016, MSE = 260.5723
Epoch 400: MAE = 13.0099, R2 = -0.0016, MSE = 260.5722
Epoch 500: MAE = 13.0099, R2 = -0.0016, MSE = 260.5722
Epoch 600: MAE = 13.0099, R2 = -0.0016, MSE = 260.5722
Epoch 700: MAE = 13.0099, R2 = -0.0016, MSE = 260.5722
Epoch 800: MAE = 13.0099, R2 = -0.0016, MSE = 260.5722
Epoch 900: MAE = 13.0099, R2 = -0.0016, MSE = 260.5723

```

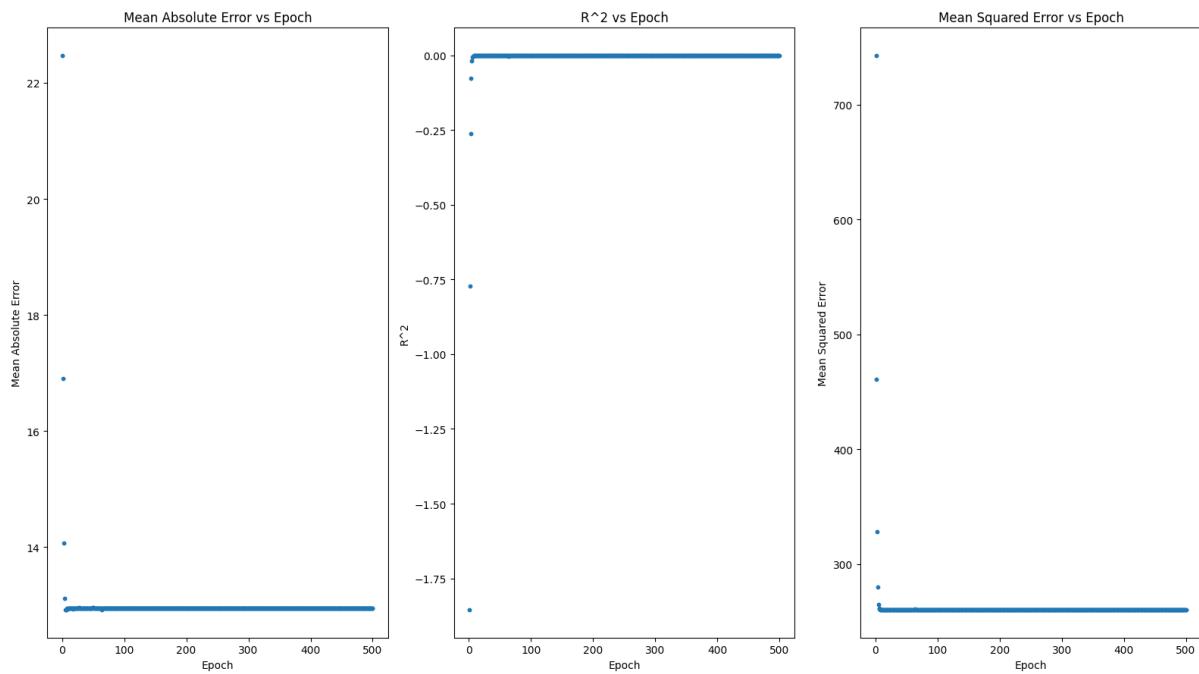


combination 18:

```

Hidden layer size: (20, 35) Activation Func: tanh Learning Rate: 0.01 Epoch/Iterations: 500
Epoch 100: MAE = 12.9452, R2 = -0.0003, MSE = 260.2476
Epoch 200: MAE = 12.9458, R2 = -0.0003, MSE = 260.2400
Epoch 300: MAE = 12.9458, R2 = -0.0003, MSE = 260.2397
Epoch 400: MAE = 12.9458, R2 = -0.0003, MSE = 260.2398

```

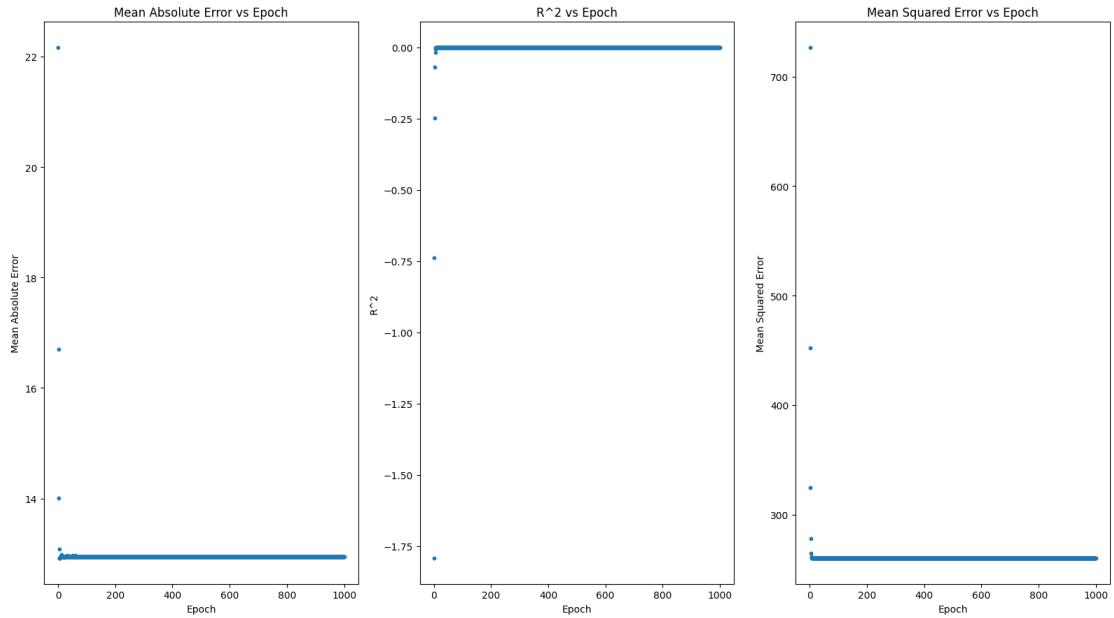


combination 19:

```

Hidden layer size: (20, 35) Activation Func: tanh Learning Rate: 0.01 Epoch/Iterations: 1000
Epoch 100: MAE = 12.9453, R2 = -0.0003, MSE = 260.2462
Epoch 200: MAE = 12.9458, R2 = -0.0003, MSE = 260.2400
Epoch 300: MAE = 12.9459, R2 = -0.0003, MSE = 260.2397
Epoch 400: MAE = 12.9458, R2 = -0.0003, MSE = 260.2398
Epoch 500: MAE = 12.9458, R2 = -0.0003, MSE = 260.2397
Epoch 600: MAE = 12.9458, R2 = -0.0003, MSE = 260.2399
Epoch 700: MAE = 12.9458, R2 = -0.0003, MSE = 260.2399
Epoch 800: MAE = 12.9458, R2 = -0.0003, MSE = 260.2399
Epoch 900: MAE = 12.9459, R2 = -0.0003, MSE = 260.2397

```

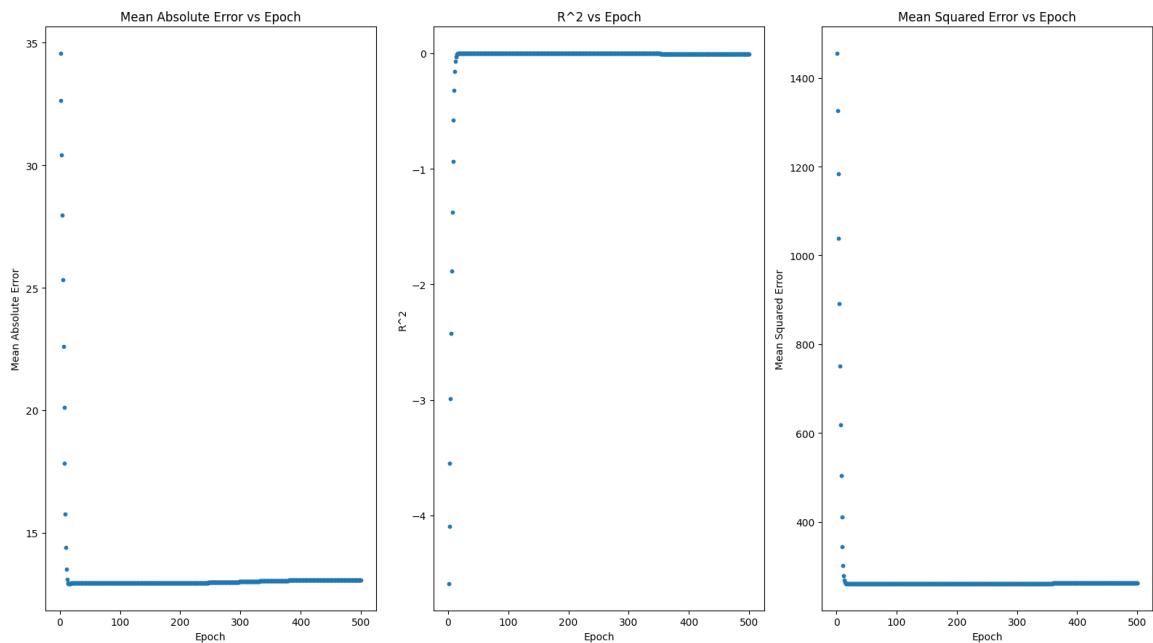


### combination 20:

```

Hidden layer size: (20, 35) Activation Func: relu Learning Rate: 0.1 Epoch/Iterations: 500
Epoch 100: MAE = 12.9439, R2 = -0.0004, MSE = 260.2628
Epoch 200: MAE = 12.9533, R2 = -0.0001, MSE = 260.1823
Epoch 300: MAE = 13.0000, R2 = -0.0011, MSE = 260.4445
Epoch 400: MAE = 13.0650, R2 = -0.0059, MSE = 261.7036

```

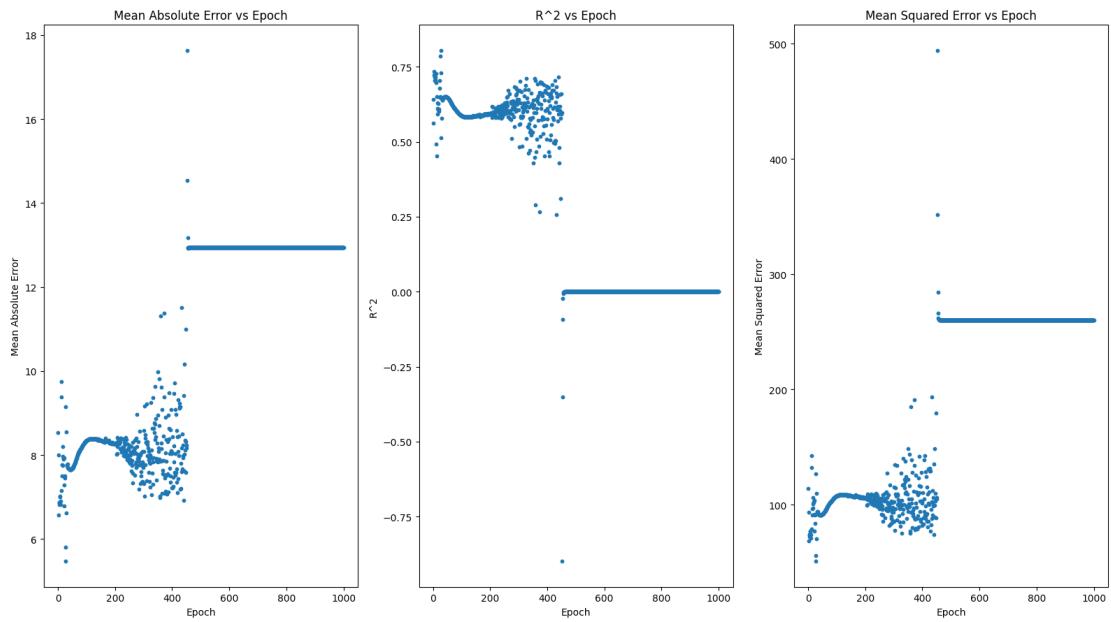


### combination 21:

```

Hidden layer size: (20, 35) Activation Func: relu Learning Rate: 0.1 Epoch/Iterations: 1000
Epoch 100: MAE = 8.3468, R2 = 0.5863, MSE = 107.6271
Epoch 200: MAE = 8.2658, R2 = 0.5937, MSE = 105.7000
Epoch 300: MAE = 8.5853, R2 = 0.5575, MSE = 115.1296
Epoch 400: MAE = 8.4302, R2 = 0.5759, MSE = 110.3479
Epoch 500: MAE = 12.9407, R2 = -0.0005, MSE = 260.3091
Epoch 600: MAE = 12.9409, R2 = -0.0005, MSE = 260.3061
Epoch 700: MAE = 12.9409, R2 = -0.0005, MSE = 260.3058
Epoch 800: MAE = 12.9409, R2 = -0.0005, MSE = 260.3058
Epoch 900: MAE = 12.9409, R2 = -0.0005, MSE = 260.3058

```

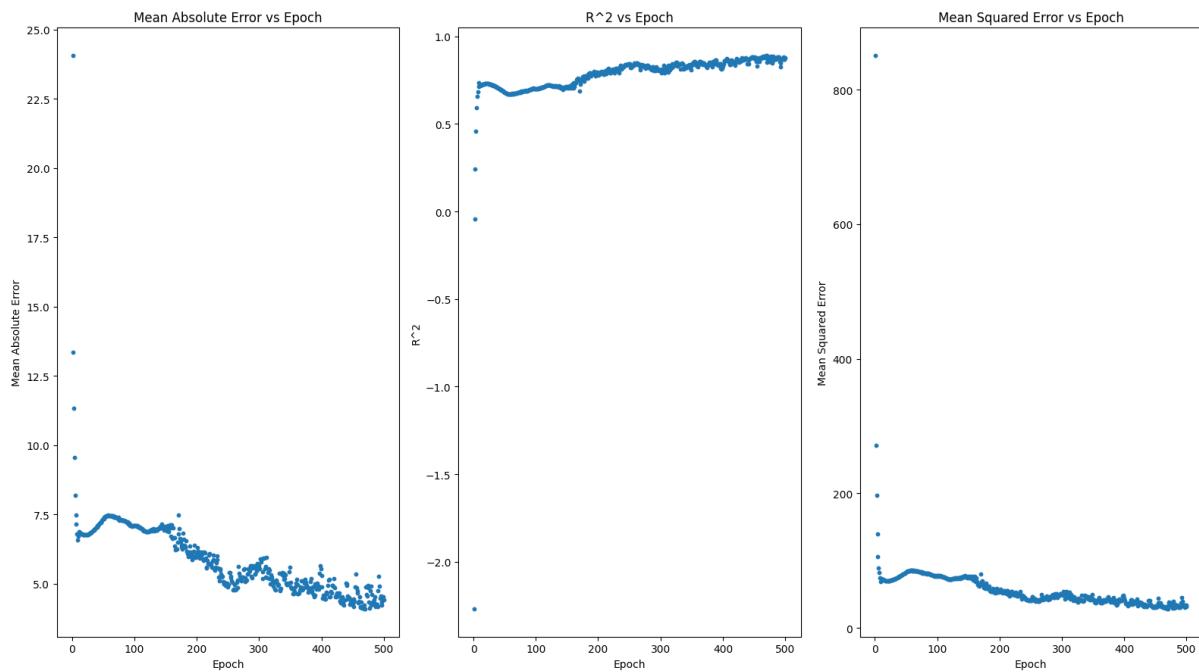


### combination 22:

```

Hidden layer size: (20, 35) Activation Func: relu Learning Rate: 0.01 Epoch/Iterations: 500
Epoch 100: MAE = 7.1012, R2 = 0.7023, MSE = 77.4409
Epoch 200: MAE = 6.3071, R2 = 0.7773, MSE = 57.9312
Epoch 300: MAE = 5.5467, R2 = 0.8121, MSE = 48.8784
Epoch 400: MAE = 4.8856, R2 = 0.8436, MSE = 40.6927

```



combination 23:

```
Hidden layer size: (20, 35) Activation Func: relu Learning Rate: 0.01 Epoch/Iterations: 1000
Epoch 100: MAE = 5.5186, R2 = 0.8042, MSE = 50.9532
Epoch 200: MAE = 5.6368, R2 = 0.8003, MSE = 51.9595
Epoch 300: MAE = 4.8560, R2 = 0.8460, MSE = 40.0587
Epoch 400: MAE = 4.3194, R2 = 0.8775, MSE = 31.8640
Epoch 500: MAE = 3.8628, R2 = 0.9000, MSE = 26.0102
Epoch 600: MAE = 3.4954, R2 = 0.9149, MSE = 22.1347
Epoch 700: MAE = 4.6913, R2 = 0.8498, MSE = 39.0649
Epoch 800: MAE = 3.8528, R2 = 0.8983, MSE = 26.4646
Epoch 900: MAE = 4.0786, R2 = 0.8856, MSE = 29.7714
```

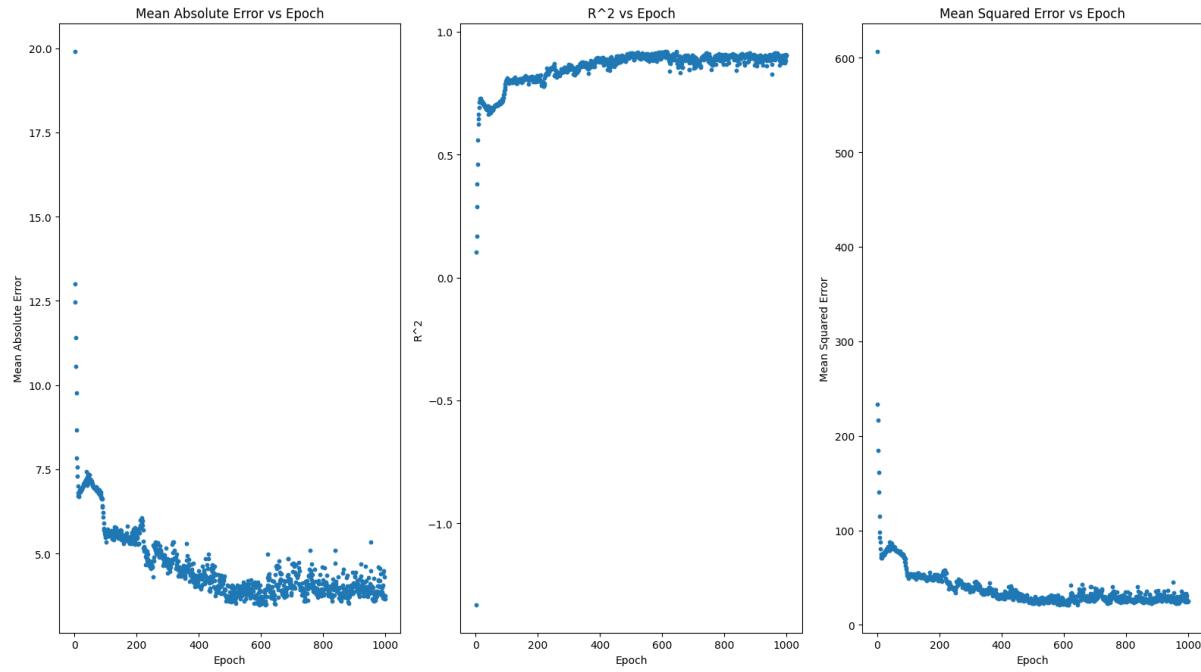


Table of combinations:

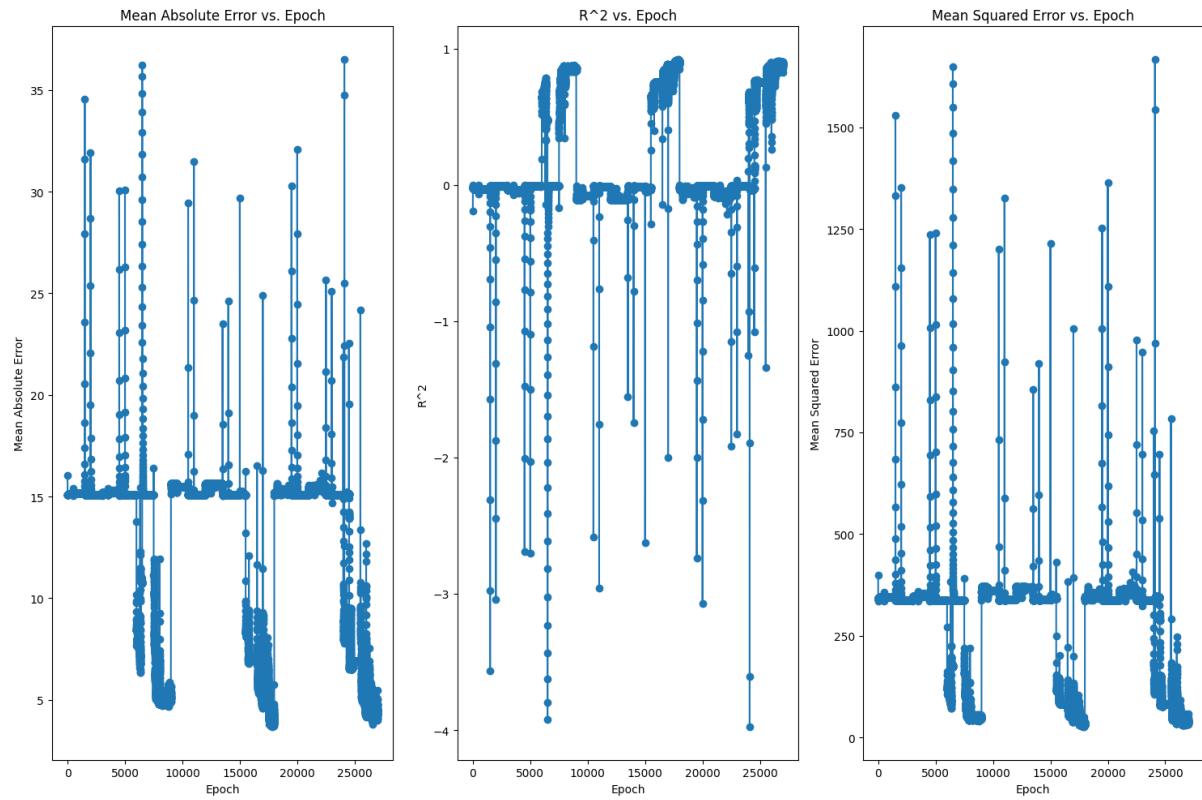
	Hidden Layer Size	Activation Func.	Learning Rate #	Epoch/Iterations
0	(10, 15)	logistic	0.10	500
1	(10, 15)	logistic	0.10	1000
2	(10, 15)	logistic	0.01	500
3	(10, 15)	logistic	0.01	1000
4	(10, 15)	tanh	0.10	500
5	(10, 15)	tanh	0.10	1000
6	(10, 15)	tanh	0.01	500
7	(10, 15)	tanh	0.01	1000
8	(10, 15)	relu	0.10	500
9	(10, 15)	relu	0.10	1000
10	(10, 15)	relu	0.01	500
11	(10, 15)	relu	0.01	1000
12	(20, 35)	logistic	0.10	500
13	(20, 35)	logistic	0.10	1000
14	(20, 35)	logistic	0.01	500
15	(20, 35)	logistic	0.01	1000
16	(20, 35)	tanh	0.10	500
17	(20, 35)	tanh	0.10	1000
18	(20, 35)	tanh	0.01	500
19	(20, 35)	tanh	0.01	1000
20	(20, 35)	relu	0.10	500
21	(20, 35)	relu	0.10	1000
22	(20, 35)	relu	0.01	500
23	(20, 35)	relu	0.01	1000

Table of model performance metric of each combination(combinations are identified by number on the table above):

	Testing Score	Training Score	MSE Testing	MSE Training	MAE Testing
0	-0.002373	-0.004793	260.784683	289.904935	13.023345
1	-0.001657	-0.003797	260.598295	289.617719	13.011721
2	-0.000458	-0.000007	260.286436	288.524128	12.942232
3	-0.000469	-0.000006	260.289331	288.523772	12.942033
4	-0.002371	-0.004790	260.784069	289.904017	13.023310
5	-0.002373	-0.004792	260.784524	289.904697	13.023336
6	-0.000472	-0.000005	260.290089	288.523685	12.941982
7	-0.000449	-0.000008	260.284094	288.524445	12.942394
8	0.708225	0.743146	75.910229	74.108114	7.027969
9	0.571848	0.624211	111.391124	108.423351	8.539773
10	0.843381	0.865344	40.747107	38.851136	4.778101
11	0.814717	0.857835	48.204607	41.017763	5.416474
12	-0.004947	-0.000856	261.454121	290.846392	13.054702
13	-0.004122	-0.007045	261.239673	290.554837	13.045814
14	-0.000278	-0.000051	260.239634	288.536759	12.945859
15	-0.000279	-0.000051	260.239659	288.536748	12.945857
16	-0.001557	-0.003653	260.572295	289.576031	13.009913
17	-0.001549	-0.003642	260.570303	289.572815	13.009772
18	-0.000279	-0.000050	260.239878	288.536650	12.945837
19	-0.000279	-0.000050	260.239889	288.536645	12.945836
20	-0.007516	-0.011087	262.122658	291.720881	13.080919
21	-0.000533	-0.000001	260.305753	288.522439	12.940950
22	0.876396	0.898603	32.157709	29.255310	4.426510
23	0.902805	0.929101	25.287053	20.456047	3.737217

	MAE Training	Mean Abs. %
0	13.802926	0.595438
1	13.788226	0.592991
2	13.689724	0.574590
3	13.689415	0.574516
4	13.802882	0.595431
5	13.802914	0.595436
6	13.689335	0.574497
7	13.689976	0.574650
8	6.659267	0.281471
9	8.185894	0.283298
10	4.687176	0.145403
11	4.846154	0.159550
12	13.845440	0.602042
13	13.832465	0.600170
14	13.695558	0.575942
15	13.695555	0.575941
16	13.785939	0.592610
17	13.785760	0.592580
18	13.695522	0.575934
19	13.695521	0.575933
20	13.881301	0.607110
21	13.687730	0.574112
22	4.043948	0.148475
23	3.497024	0.124478

### Model history with all the combinations combined:



### Final conclusion/Best Parameters:

From the trials of all combinations, we could conclude that our best set of parameters are:

Hidden layer size: (20,35)

Activation function: Relu function

Learning rate: 0.01

Epoch: 1000

Data Split: 80% training 20% testing

Outlier threshold for preprocessed data: 1.5

Combination/Trial 23 had the most consistent and high  $R^2$  values/accuracy and the lowest MSE, MAE values among all combinations. This tells us that this combination of hyper-parameter values is the best combination to accurately predict the target variable  $y$ .

In regards to our dataset, the ReLu activation function had the best performance regardless of other parameters. Out performing the other activation functions with whatever combination of other hyper-parameters given.

