Assignment 3

Min-Wei,Li

Data 1

weather_izmir

- Data_id = 42369
- The file contains the weather information of Izmir from 01/01/1994 to 31/12/1997. From given features, the goal is to predict the mean temperature.
- There are no nominal features and missing values , both target and features are all numeric , so I didn’t transform the data
- https://www.openml.org/search?type=data&sort=qualities.NumberOfNumericFeatures&status=active&qualities.NumberOfClasses=lte_1&qualities.NumberOfInstances=between_1000_10000&order=asc&qualities.NumberOfFeatures=between_10_100&id=42369

Data 1 information

	Mean_temperature (target)	numeric	489 distinct values 0 missing attributes
	Max_temperature	numeric	300 distinct values 0 missing attributes
	Min_temperature	numeric	284 distinct values 0 missing attributes
	Dewpoint	numeric	387 distinct values 0 missing attributes
	Precipitation	numeric	26 distinct values 0 missing attributes
	Sea_level_pressure	numeric	90 distinct values 0 missing attributes
	Standard_pressure	numeric	52 distinct values 0 missing attributes

Model 1

Input Layer:

The model has 9 input nodes, representing the dataset's 10 features, excluding the target variable.

Hidden Layer:

Contains 1 Dense layer with 5 nodes.

Activation function: Sigmoid.

Parameters: 50 (calculated as 9 input features $\times$ 5 nodes + 5 biases).

Output Layer:

Contains 1 Dense layer with 1 node.

Designed for regression tasks.

Parameters: 6 (calculated as 5 hidden nodes $\times$ 1 output node + 1 bias).

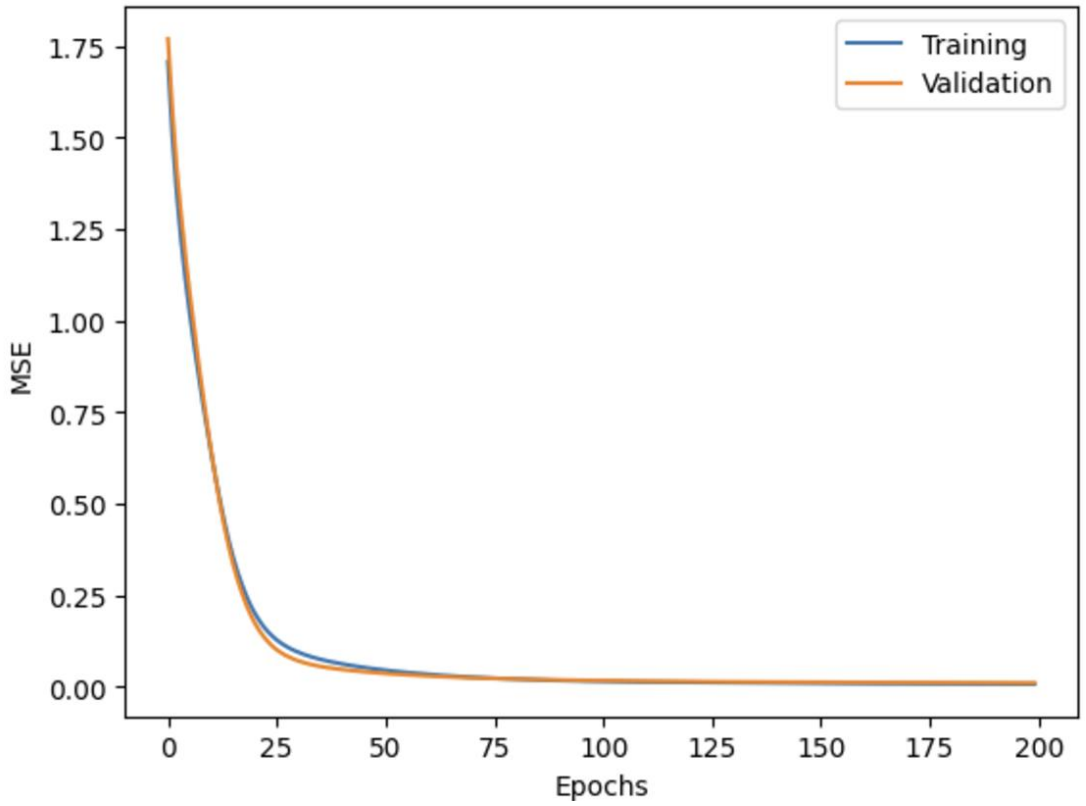
Model 1 summary

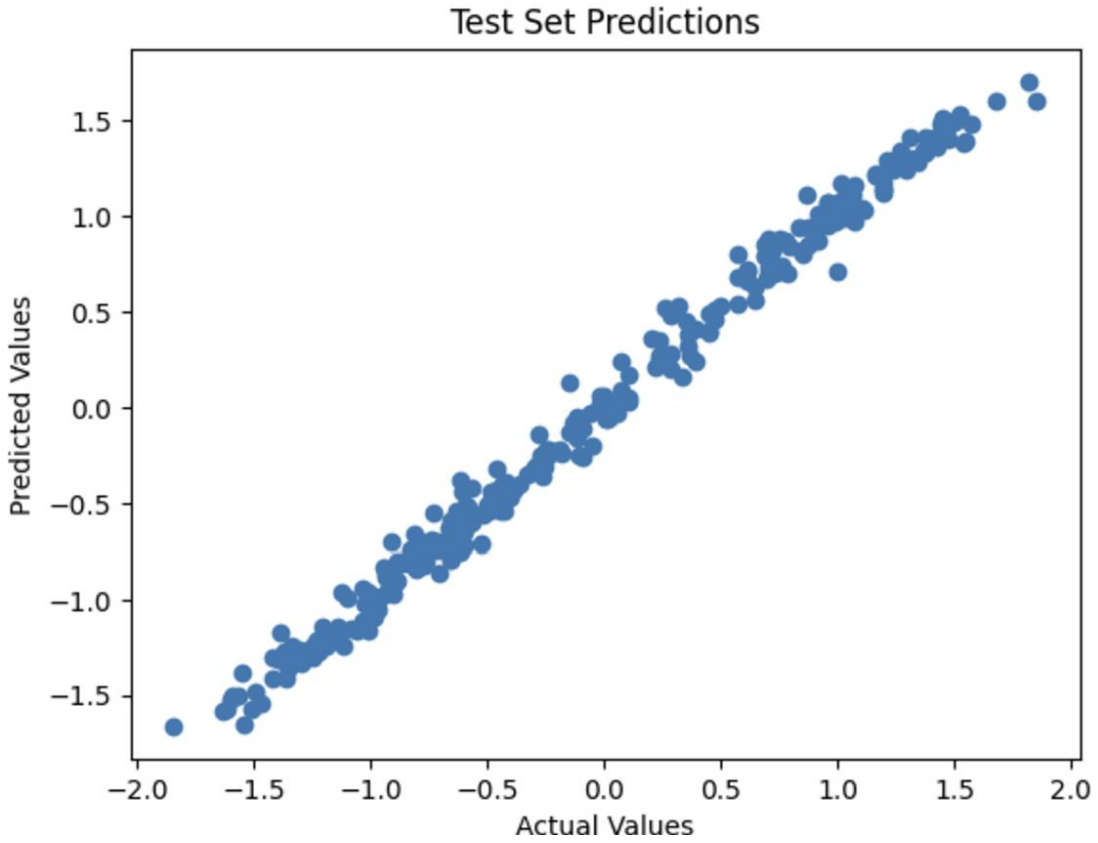
Model: "sequential_21"

Layer (type)	Output Shape	Param #
dense_40 (Dense)	(None, 5)	50
dense_41 (Dense)	(None, 1)	6

Total params: 56 (224.00 B)
Trainable params: 56 (224.00 B)
Non-trainable params: 0 (0.00 B)

Model 1 Result





Model 2

Input Layer:

The model has 9 input nodes, representing the dataset's 10 features, excluding the target variable.

Hidden Layer:

Contains 1 Dense layer with 30 nodes.

Activation function: Sigmoid.

Parameters: 300 (calculated as 9 input features $\times$ 30 nodes + 30 biases).

Output Layer:

Contains 1 Dense layer with 1 node.

Designed for regression tasks.

Parameters: 31 (calculated as 30 hidden nodes $\times$ 1 output node + 1 bias).

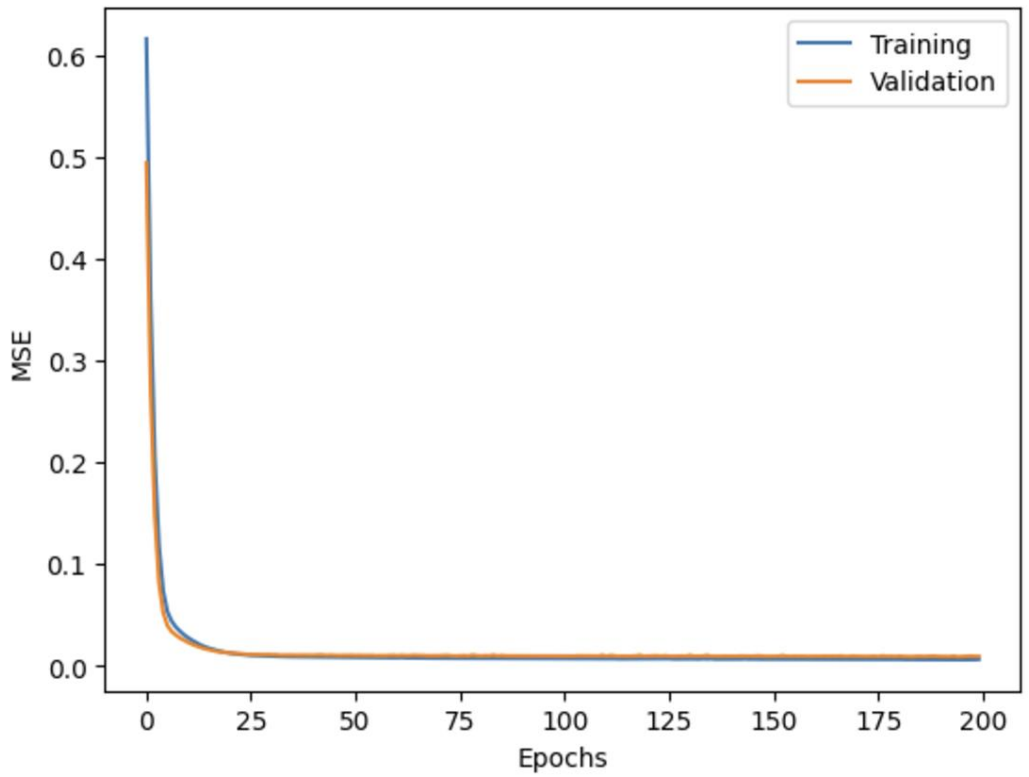
Model 2 summary

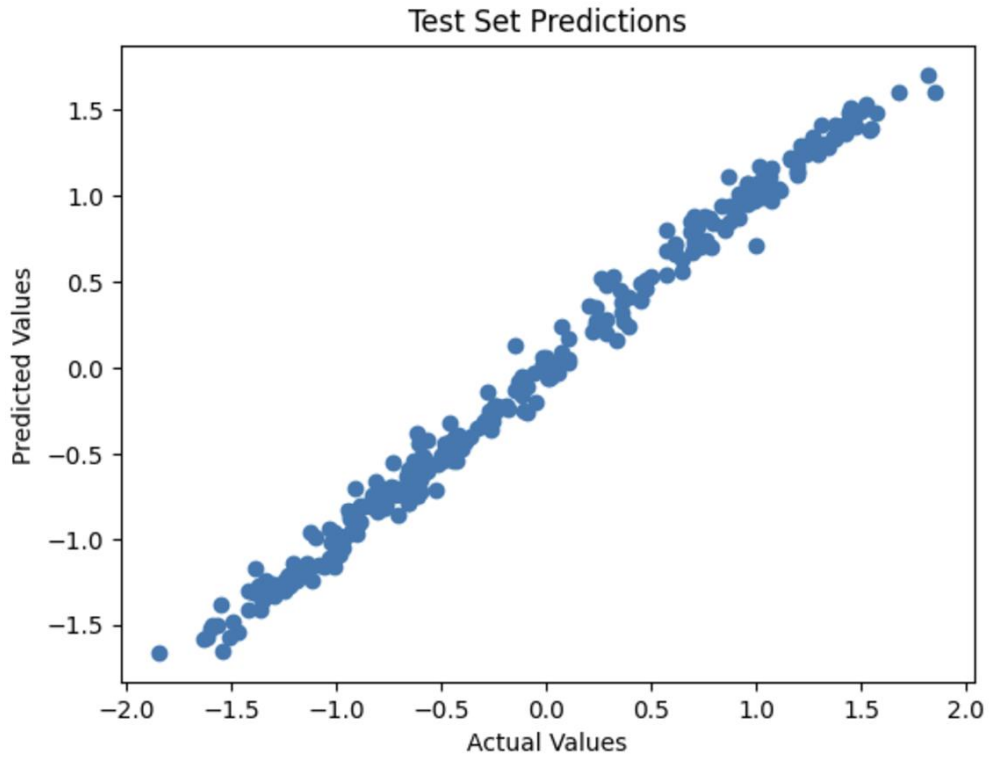
Model: "sequential_20"

Layer (type)	Output Shape	Param #
dense_38 (Dense)	(None, 30)	300
dense_39 (Dense)	(None, 1)	31

Total params: 331 (1.29 KB)
Trainable params: 331 (1.29 KB)
Non-trainable params: 0 (0.00 B)

Model 2 Result





Model 3

Input Layer:

The model has 9 input nodes, representing the dataset's 10 features, excluding the target variable.

Hidden Layer:

Contains 1 Dense layer with 200 nodes.

Activation function: Sigmoid.

Parameters: 2,000 (calculated as 9 input features $\times$ 200 nodes + 200 biases).

Output Layer:

Contains 1 Dense layer with 1 node.

Designed for regression tasks.

Parameters: 201 (calculated as 200 hidden nodes $\times$ 1 output node + 1 bias).

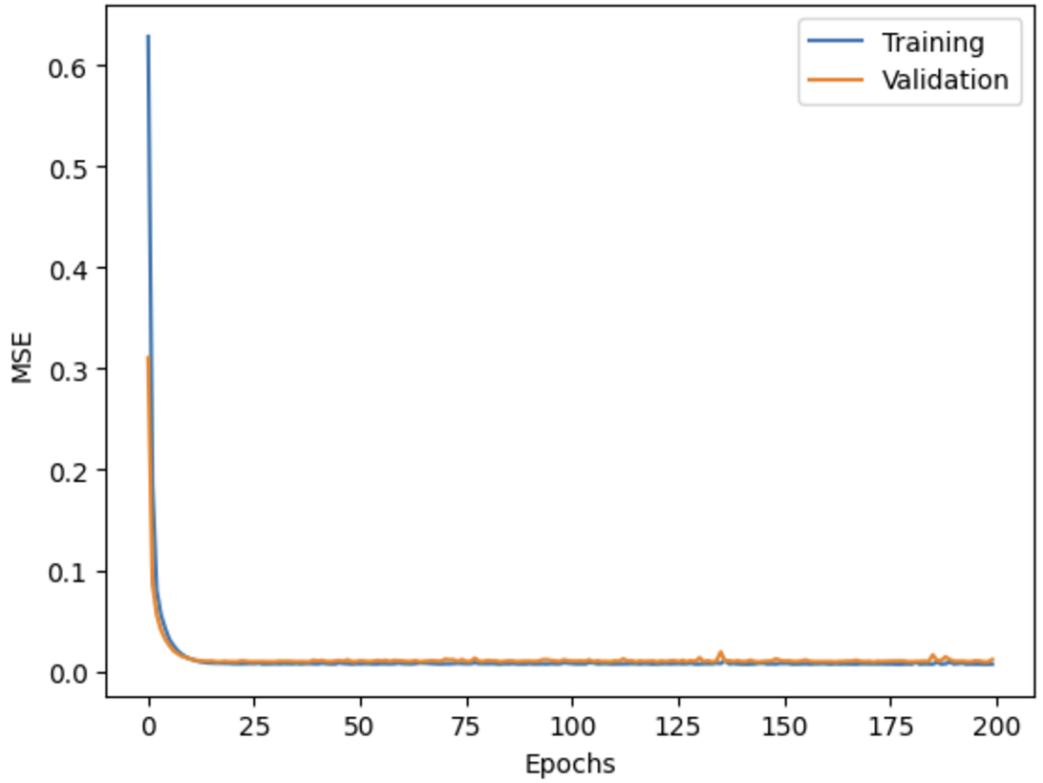
Model 3 summary

Model: "sequential_19"

Layer (type)	Output Shape	Param #
dense_36 (Dense)	(None, 200)	2,000
dense_37 (Dense)	(None, 1)	201

Total params: 2,201 (8.60 KB)
Trainable params: 2,201 (8.60 KB)
Non-trainable params: 0 (0.00 B)

Model 3 Result



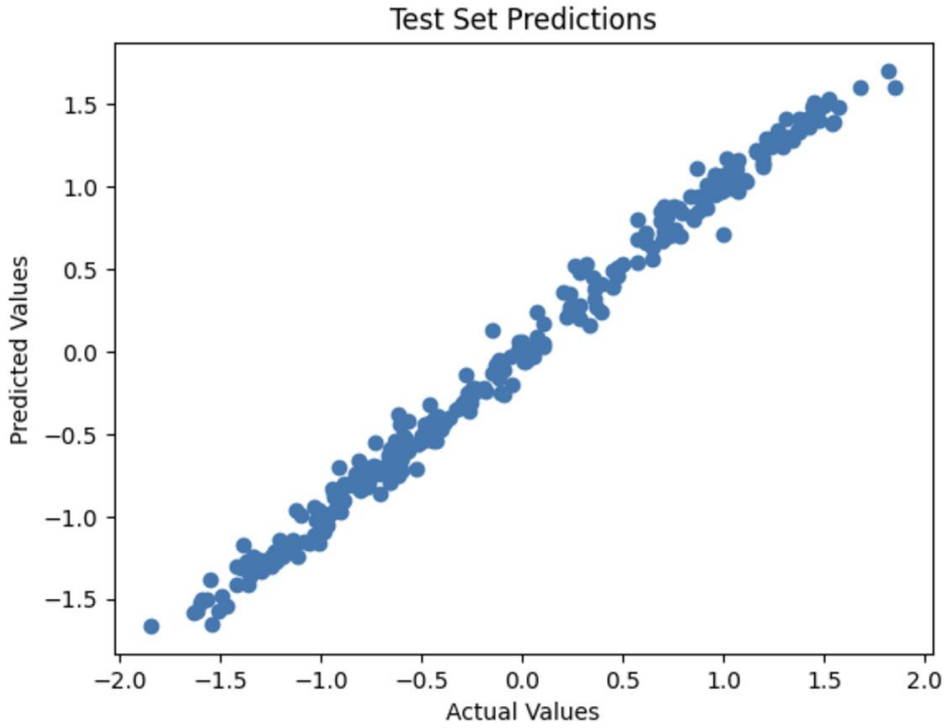


table of test errors

	Model	Test MSE
0	Model 1 (5 nodes)	0.007141
1	Model 2 (30 nodes)	0.006856
2	Model 3 (200 nodes)	0.007815

Conclusion

- **Training and Validation Errors:**

All three models show a steady decline in both training and validation MSE across epochs.

The validation error closely follows the training error, indicating that the models are not overfitting.

- **Comparison of Test MSE:**

Model 1 (5 nodes): Test MSE = 0.007141

Model 2 (30 nodes): Test MSE = 0.006856

Model 3 (200 nodes): Test MSE = 0.007815

Model 2 achieves the lowest test error, suggesting that 30 nodes in the hidden layer offer the best balance of model complexity and generalization.

- **Model Complexity and Performance:**

Model 1, with only 5 nodes, may be underfitting the data slightly, as its test error is higher than Model 2.

Model 3, with 200 nodes, shows a slight increase in test error compared to Model 2, potentially due to overfitting, as the model is too complex for the dataset.

- **Optimal Model:**

Model 2 strikes the best balance between complexity and performance, achieving the lowest test MSE.

Adding too many nodes (Model 3) does not improve performance and may increase the risk of overfitting.

Data 2

wind

- Data_id = 503
- wind daily average wind speeds for 1961-1978 at 12 synoptic meteorological stations in the Republic of Ireland (Haslett and raftery 1989).
- There are no nominal features and missing values , both target and features are all numeric , so I didn’t transform the data

- https://www.openml.org/search?type=data&sort=runs&status=active&qualities.NumberOfInstances=between_1000_10000&qualities.NumberOfFeatures=between_10_100&qualities.NumberOfClasses=lte_1&id=503

Data 2 information

15 Features

▾ Expand

	Feature Name	Type	Distinct/Missing Values	Ontology
	MAL (target)	numeric	779 distinct values 0 missing attributes	
	year	numeric	18 distinct values 0 missing attributes	
	month	numeric	12 distinct values 0 missing attributes	
	day	numeric	31 distinct values 0 missing attributes	
	RPT	numeric	671 distinct values 0 missing attributes	
	VAL	numeric	607 distinct values 0 missing attributes	
	ROS	numeric	611 distinct values 0 missing attributes	

Model 1

Input Layer:
The model has 14 input nodes, representing the dataset's 15 features, excluding the target variable.

Hidden Layer:
Contains 1 Dense layer with 5 nodes.
Activation function: Sigmoid.
Parameters: 75 (calculated as 14 input features×5 nodes+5 biases14 input features×5 nodes+5 biases).

Output Layer:
Contains 1 Dense layer with 1 node.
Designed for regression tasks.
Parameters: 6 (calculated as 5 hidden nodes×1 output node+1 bias5 hidden nodes×1 output node+1 bias)

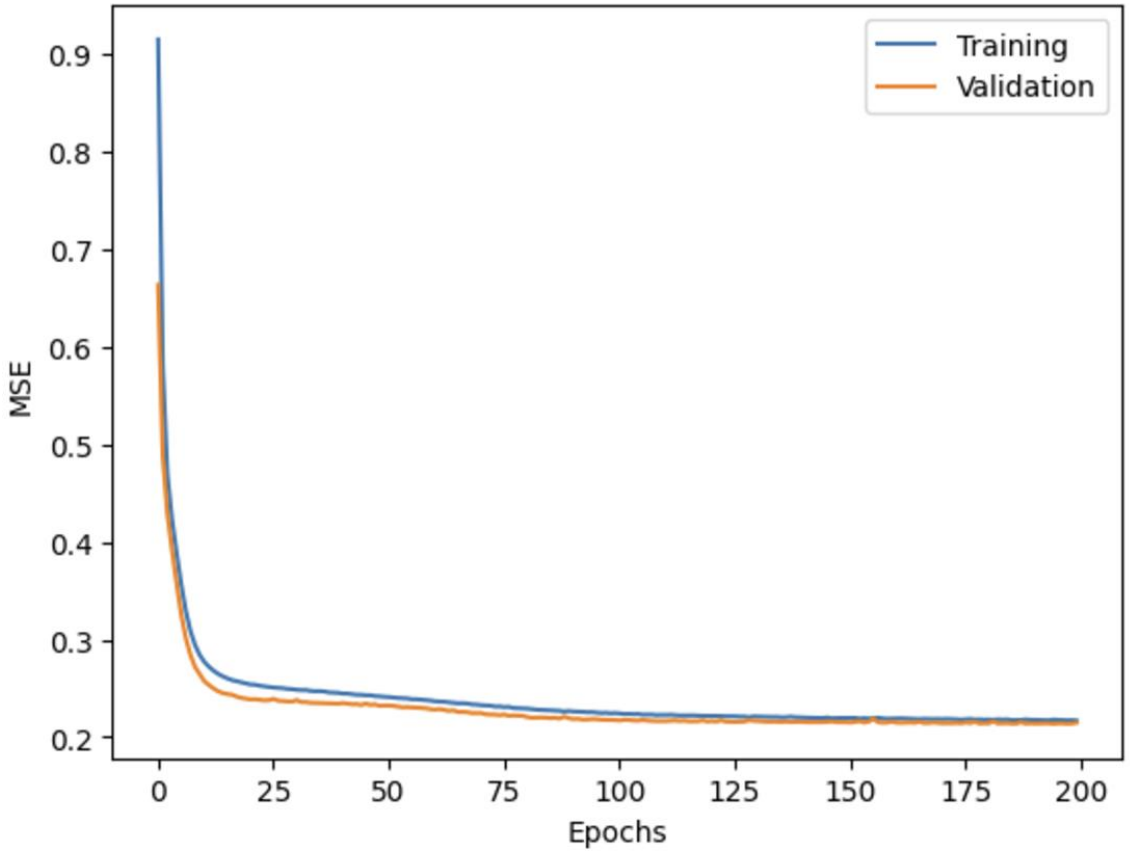
Model 1 summary

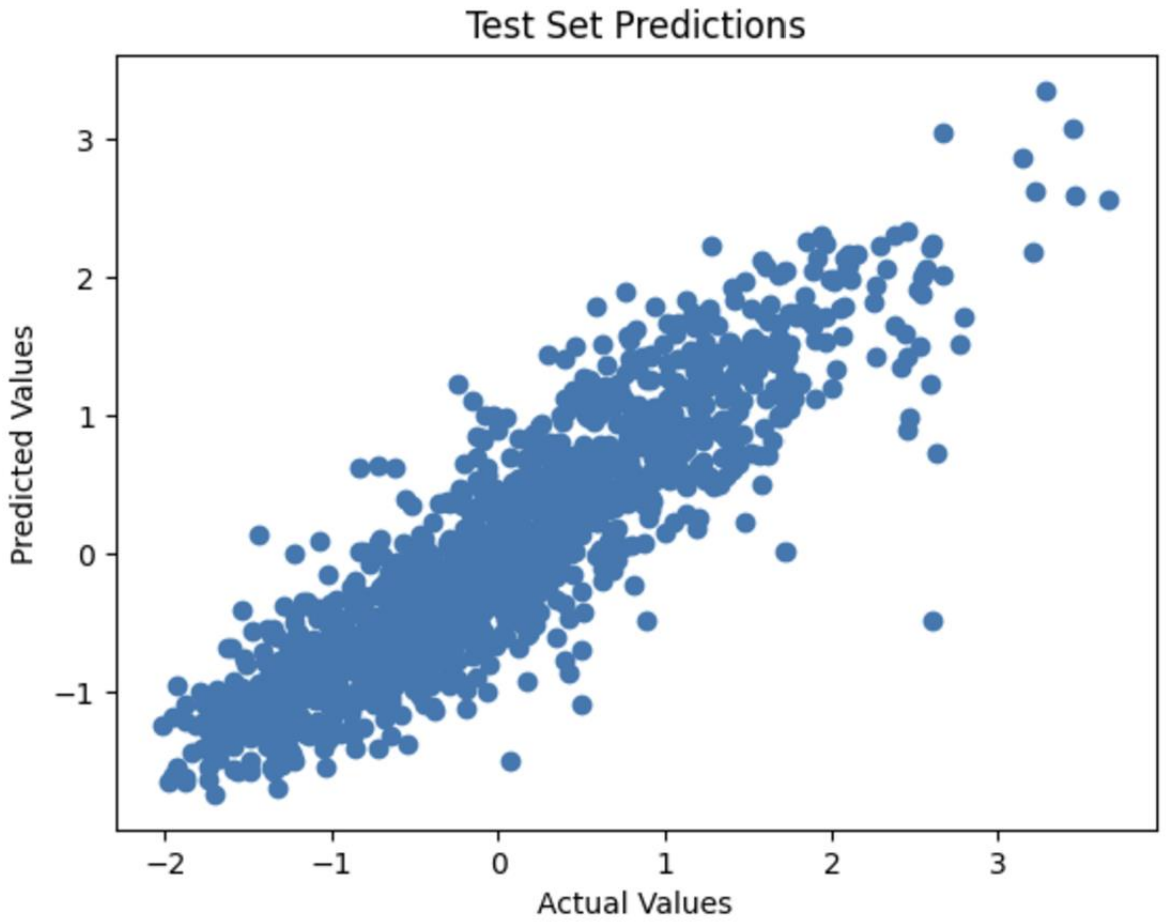
Model: "sequential_22"

Layer (type)	Output Shape	Param #
dense_42 (Dense)	(None, 5)	75
dense_43 (Dense)	(None, 1)	6

Total params: 81 (324.00 B)
Trainable params: 81 (324.00 B)
Non-trainable params: 0 (0.00 B)

Model 1 Result





Model 2

Input Layer:
The model has 14 input nodes, representing the dataset's 15 features, excluding the target variable.

Hidden Layer:
Contains 1 Dense layer with 30 nodes.
Activation function: Sigmoid.
Parameters: 300 (calculated as 14 input features×30 nodes+30 biases14 input features×30 nodes+30 biases).

Output Layer:
Contains 1 Dense layer with 1 node.
Designed for regression tasks.
Parameters: 31 (calculated as 30 hidden nodes×1 output node+1 bias30 hidden nodes×1 output node+1 bias).

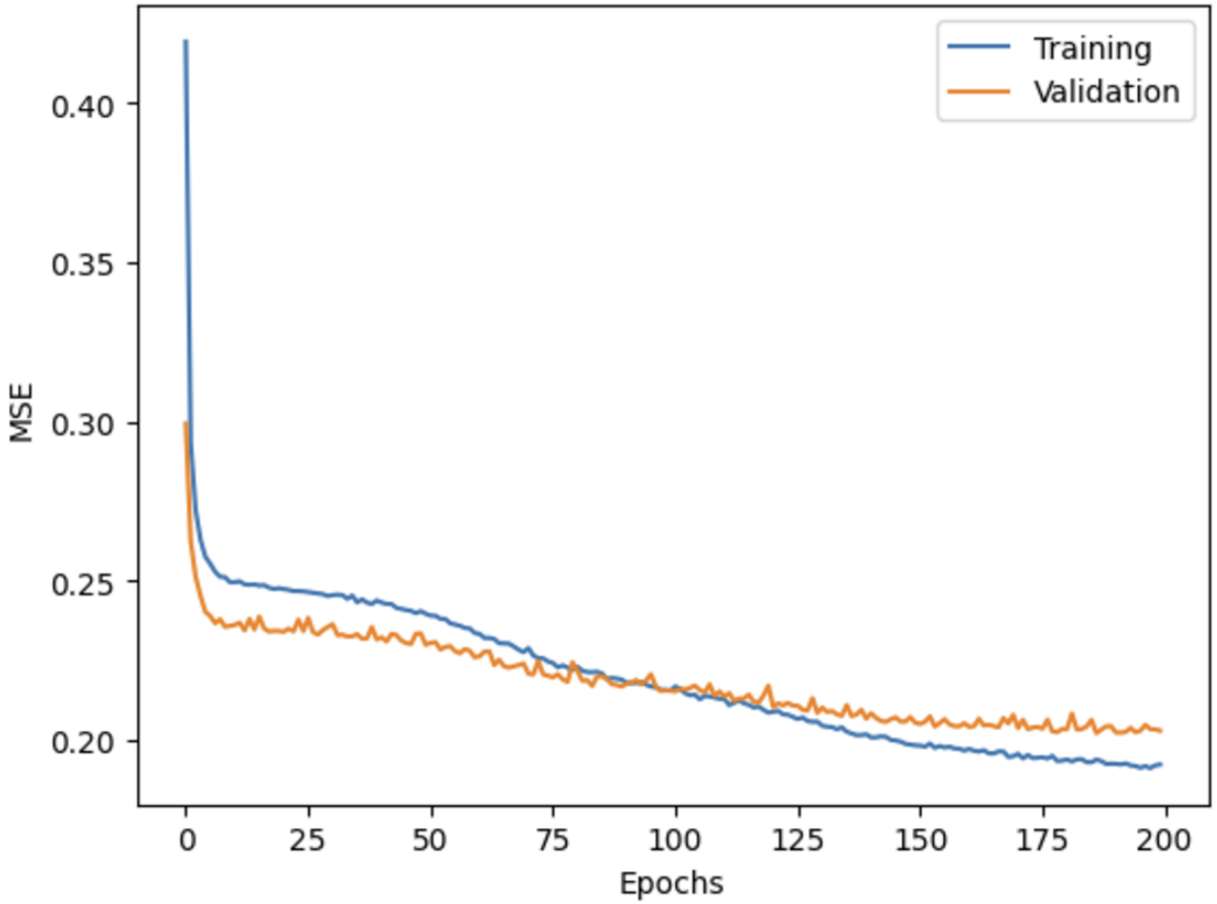
Model 2 summary

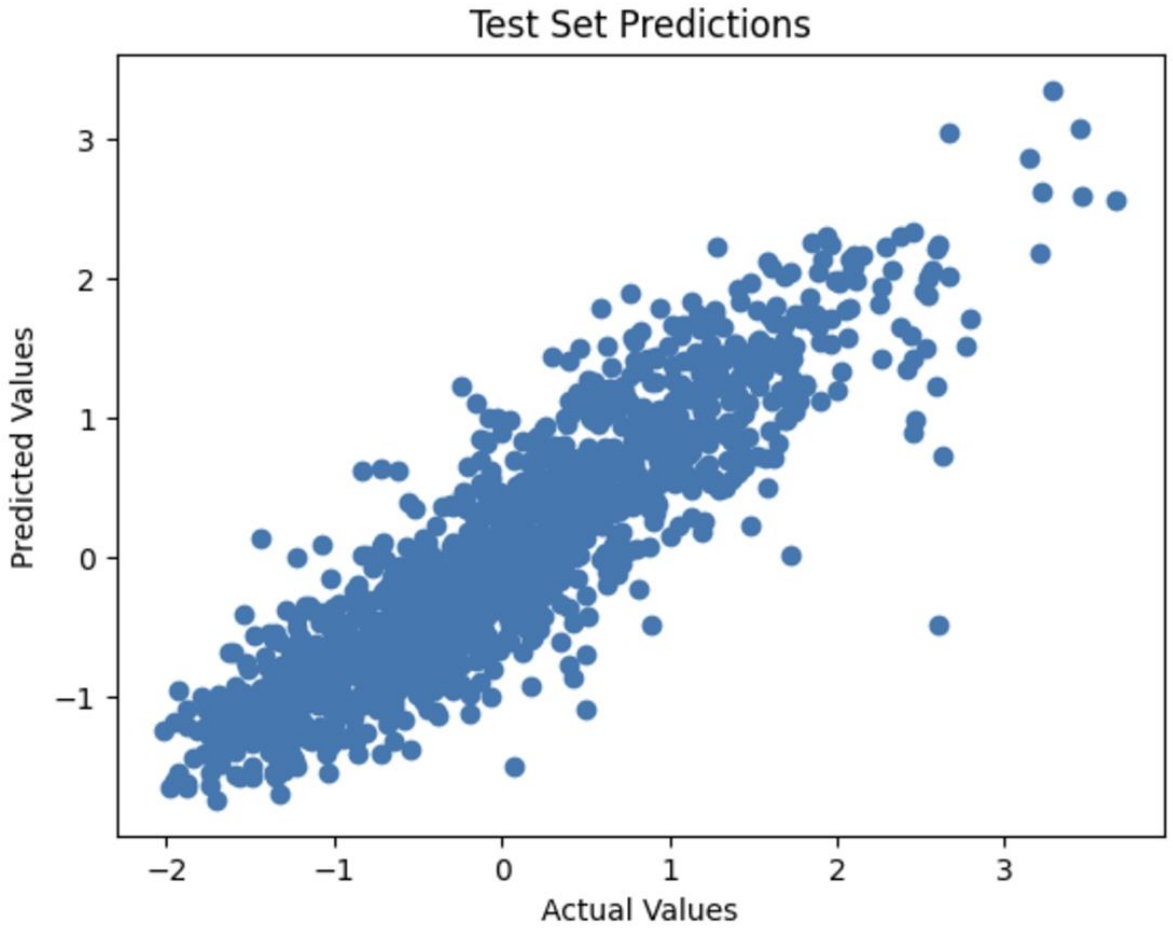
Model: "sequential_20"

Layer (type)	Output Shape	Param #
dense_38 (Dense)	(None, 30)	300
dense_39 (Dense)	(None, 1)	31

Total params: 331 (1.29 KB)
Trainable params: 331 (1.29 KB)
Non-trainable params: 0 (0.00 B)

Model 2 Result





Model 3

Input Layer:

The model has 14 input nodes, representing the dataset's 15 features, excluding the target variable.

Hidden Layer:

Contains 1 Dense layer with 200 nodes.

Activation function: Sigmoid.

Parameters: 3,000 (calculated as
14 input features×200 nodes+200 biases14 input features×200 nodes+200 biases).

Output Layer:

Contains 1 Dense layer with 1 node.

Designed for regression tasks.

Parameters: 201 (calculated as 200 hidden nodes×1 output node+1 bias200 hidden nodes×1 output node+1 bias).

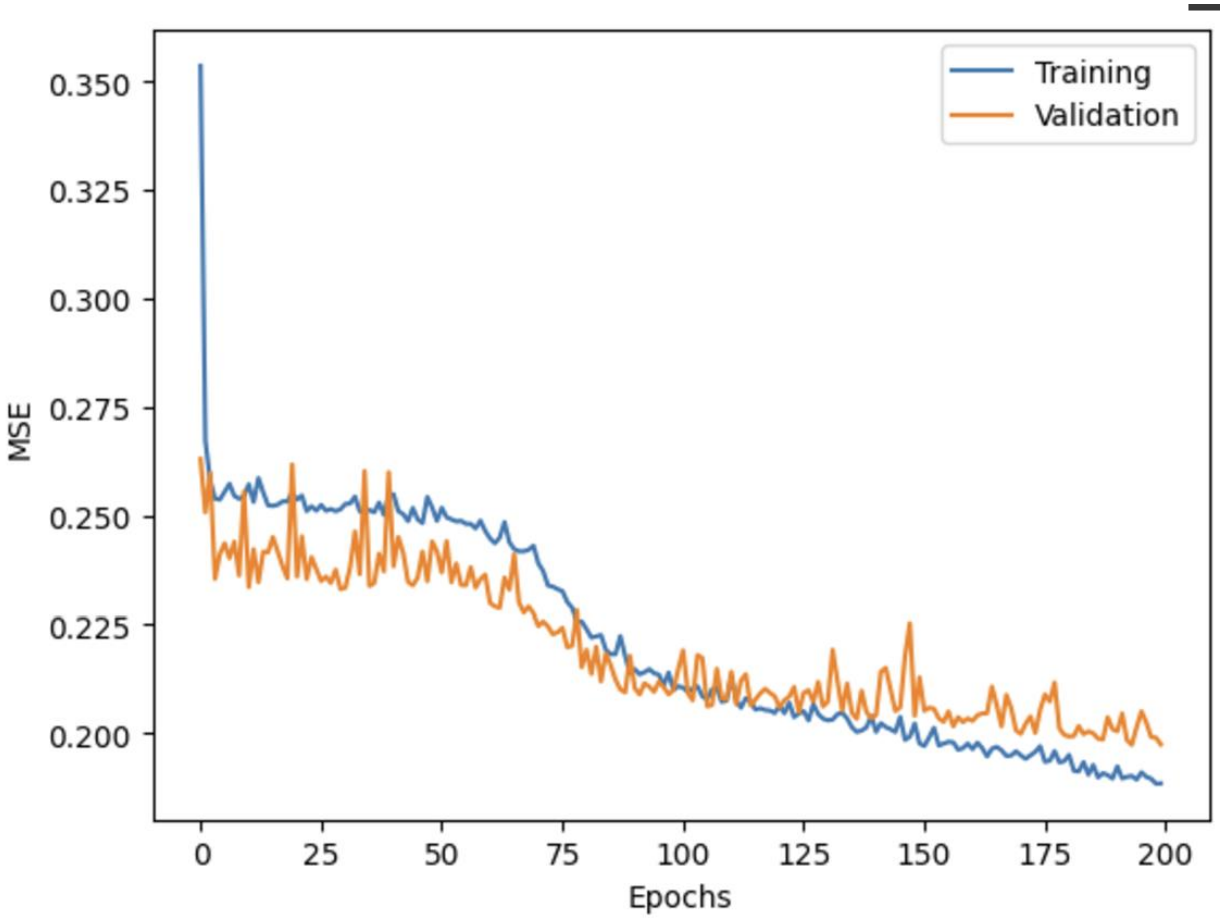
Model 3 summary

Model: "sequential_24"

Layer (type)	Output Shape	Param #
dense_46 (Dense)	(None, 200)	3,000
dense_47 (Dense)	(None, 1)	201

Total params: 3,201 (12.50 KB)
Trainable params: 3,201 (12.50 KB)
Non-trainable params: 0 (0.00 B)

Model 3 Result



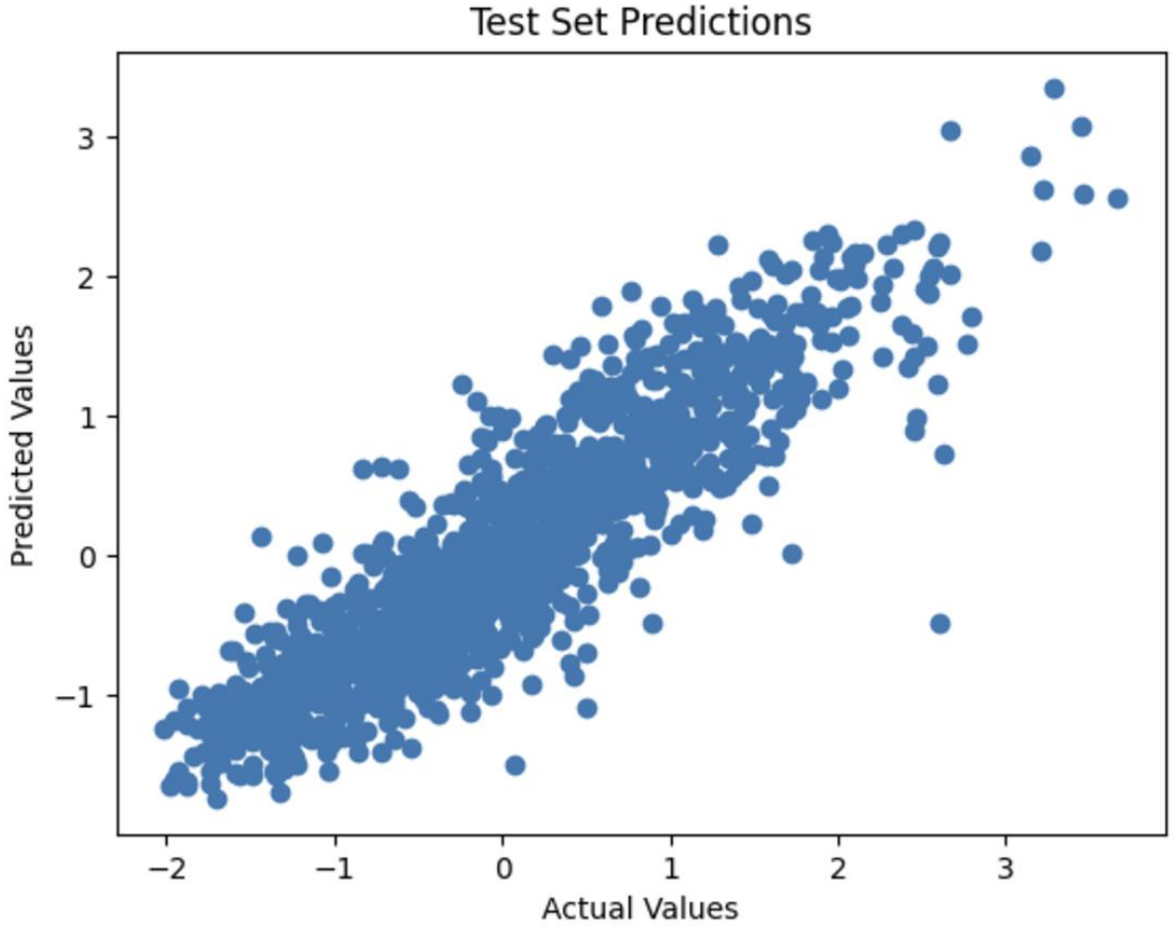


table of test errors

	Model	Test MSE
0	Model 1 (5 nodes)	0.207440
1	Model 2 (30 nodes)	0.188445
2	Model 3 (200 nodes)	0.194654

Conclusion

- **Training and Validation Errors:**

All three models show a steady decline in both training and validation MSE across epochs.

The validation error closely follows the training error, indicating that the models are not overfitting.

- **Comparison of Test MSE:**

Model 1 (5 nodes): Test MSE = 0.007141

Model 2 (30 nodes): Test MSE = 0.006856

Model 3 (200 nodes): Test MSE = 0.007815

Model 2 achieves the lowest test error, suggesting that 30 nodes in the hidden layer offer the best balance of model complexity and generalization.

- **Model Complexity and Performance:**

Model 1, with only 5 nodes, may be underfitting the data slightly, as its test error is higher than Model 2.

Model 3, with 200 nodes, shows a slight increase in test error compared to Model 2, potentially due to overfitting, as the model is too complex for the dataset.

- **Optimal Model:**

Model 2 strikes the best balance between complexity and performance, achieving the lowest test MSE.

Adding too many nodes (Model 3) does not improve performance and may increase the risk of overfitting.