Data Science 2

Project 2

By: Rohith Lingala,
Harshith Kasthuri,
Charan Reddy Mannuru,
James Notoma,
Josh Kolesar

UNIVERSITY OF GEORGIA

Auto MPG Best Performances (2L)

Performance (Python)

2-Layer Neural Network:
Best R^2: 0.9277
Activation and Optimizer: Activation: Tanh, Optimizer: SGD
Metrics:
In-Sample MSE: 6.1132707595825195
In-Sample RMSE: 2.4725029468536377
In-Sample R^2: 0.9025
Validation MSE: 4.470783233642578
Validation RMSE: 2.114422559738159
Validation R^2: 0.9168
Cross-Validation MSE: 4.506539821624756
Cross-Validation RMSE: 2.0948846340179443
Cross-Validation R^2: 0.9277

Performance (Scala)

- Sigmoid, eta:0.1, epochs: 400
  - R-Squared: 0.863
  - Adjusted R-Squared: 0.860
  - MSE: 8.34263
  - RMSE: 2.88836
  - MAE: 2.15410
  - SSE: 3320.37

Auto MPG Best Performances (3L)

Performance (Python)

3-Layer Neural Network:
Best R^2: 0.9568
Activation and Optimizer: Activation: ReLU, Optimizer: SGD
Metrics:
In-Sample MSE: 5.085403919219971
In-Sample RMSE: 2.2550840377807617
In-Sample R^2: 0.9189
Validation MSE: 4.543299198150635
Validation RMSE: 2.1315016746520996
Validation R^2: 0.9155
Cross-Validation MSE: 2.7511565685272217
Cross-Validation RMSE: 1.5926520824432373
Cross-Validation R^2: 0.9568

Performance (Scala)

- tanh, id, eta: 0.1, epochs:400
  - R-Squared: 0.948
  - Adjusted R-Squared: 0.947
  - MSE: 3.15687
  - RMSE: 1.77676
  - MAE: 1.31662
  - SSE: 1256.43

Auto MPG Best Performances (XL)

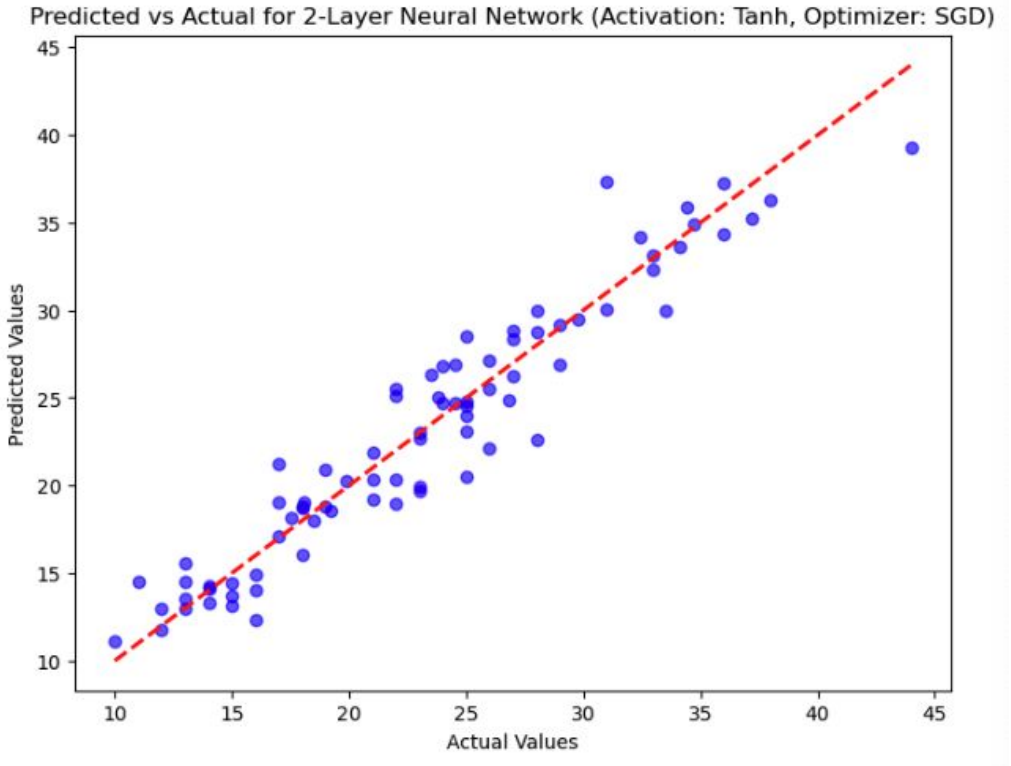
Performance (Python)

XL Neural Network:
- Best R^2: 0.9818
- Activation and Optimizer: Activation: ReLU, Optimizer: SGD
- Metrics:
  - In-Sample MSE: 3.450523853302002
  - In-Sample RMSE: 1.8575586080551147
  - In-Sample R^2: 0.9450
  - Validation MSE: 5.17495059967041
  - Validation RMSE: 2.2748517990112305
  - Validation R^2: 0.9038
  - Cross-Validation MSE: 1.1483126878738403
  - Cross-Validation RMSE: 1.0151135921478271
  - Cross-Validation R^2: 0.9818

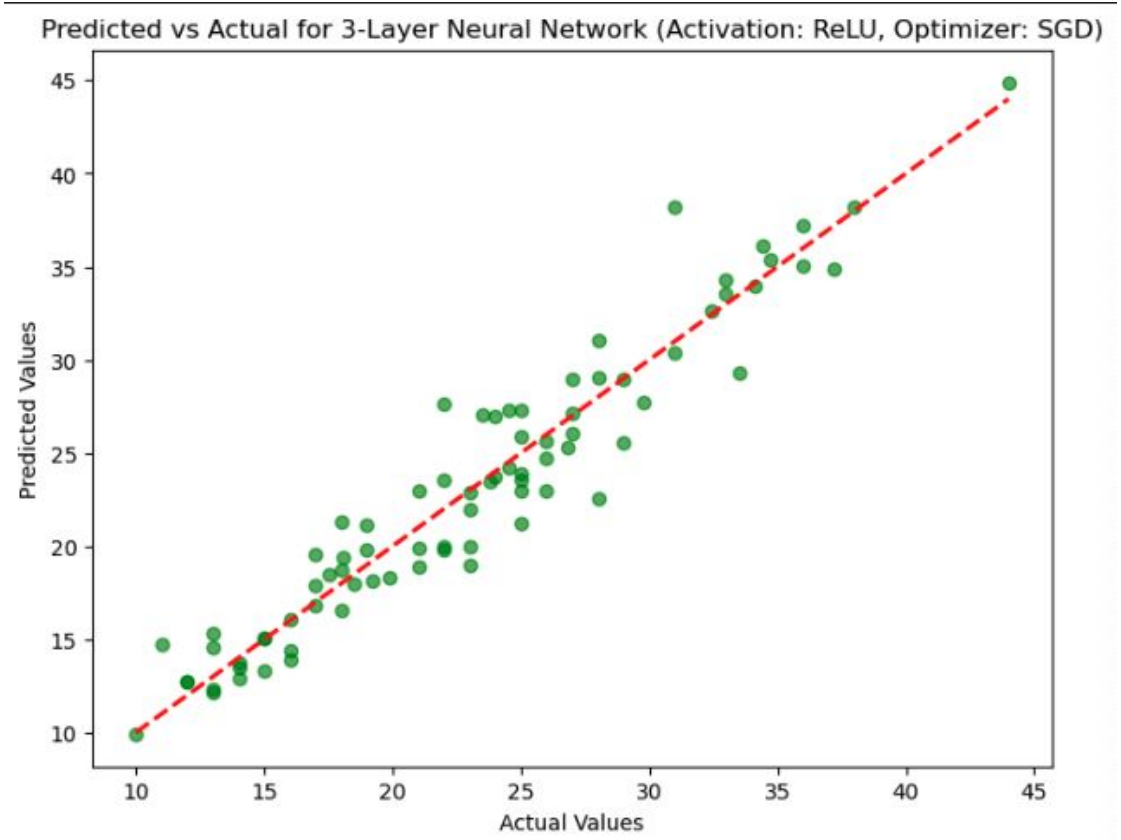
Performance (Scala)

- Tanh, tanh, id, eta: 0.1, epochs:400
  - R-Squared: 0.9707
  - Adjusted R-Squared: 0.9702
  - MSE: 1.78208
  - RMSE: 1.33495
  - MAE: 0.974232
  - SSE: 709.268

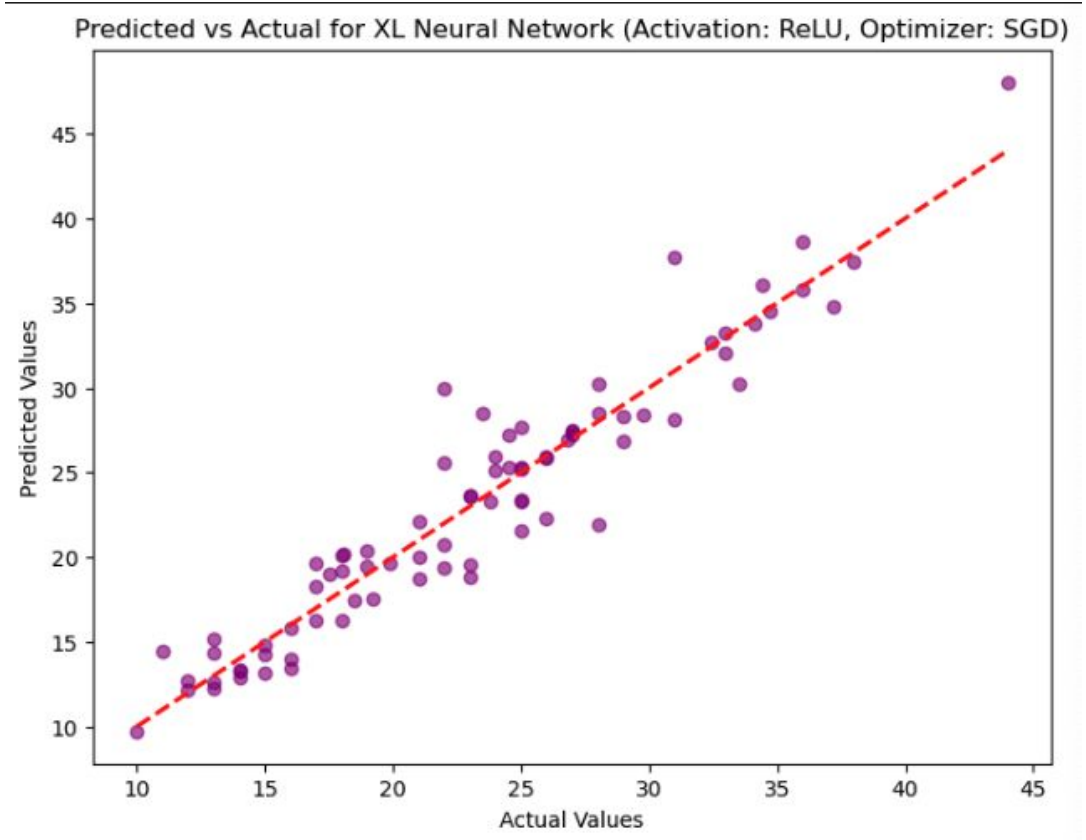
Auto MPG Plots (Python, 2L)



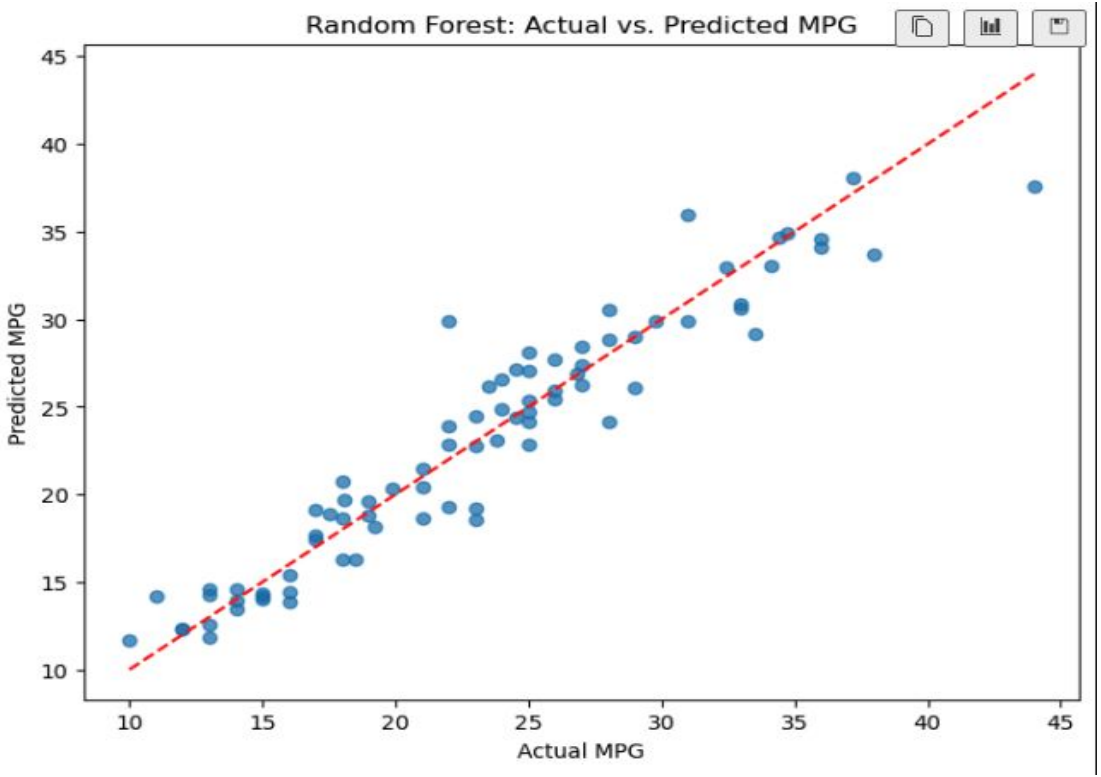
Auto MPG Plots (Python, 3L)



Auto MPG Plots (Python, XL)



AutoMPG Python(Random Forest Regressor)

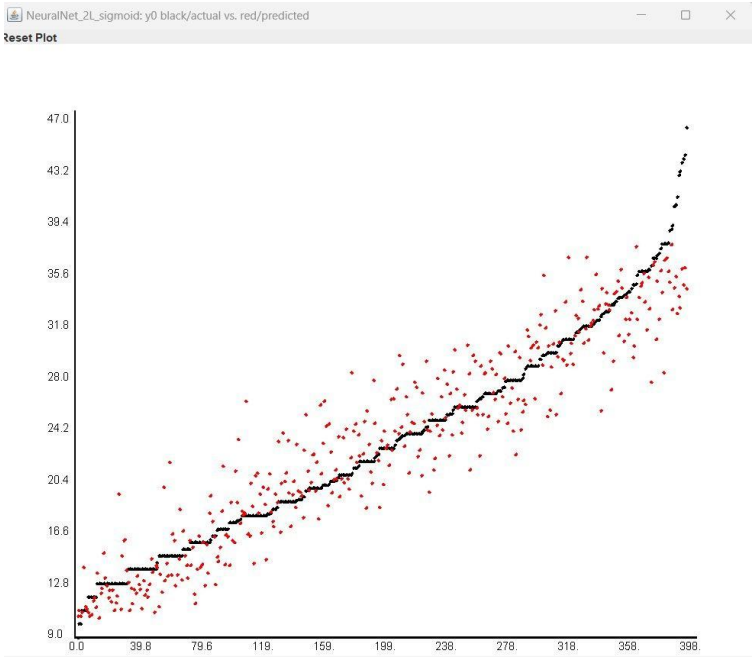


Python:
Random Forest Regressor:

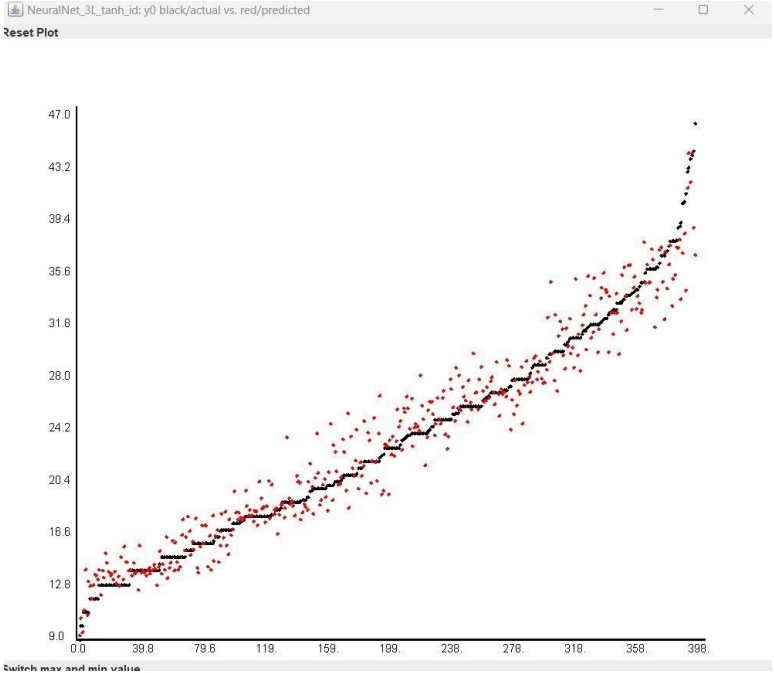
R2:0.9160.

Auto MPG Plots (Scala)

Neural Net 2L

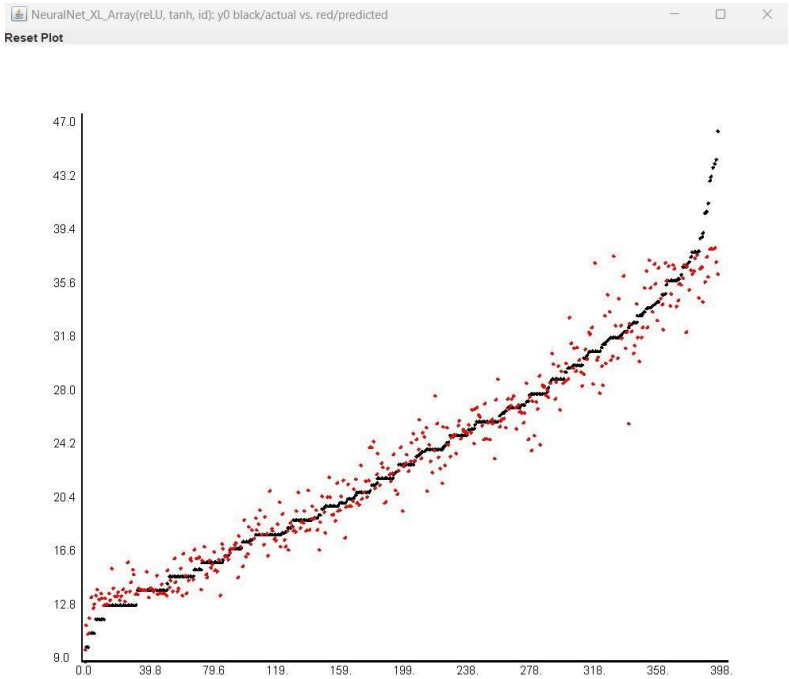


Neural Net 3L

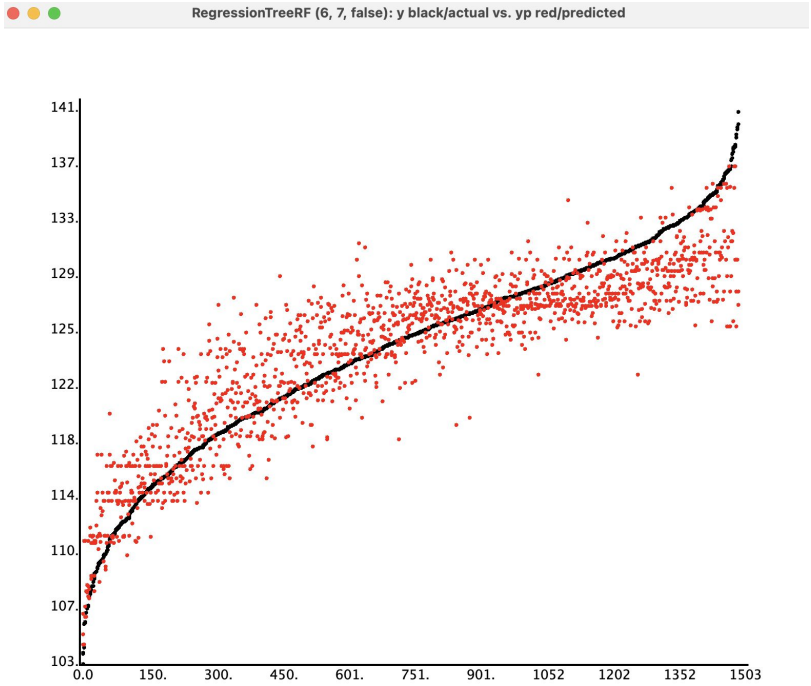


Auto MPG Plots (Scala)

Neural Net XL



Random Forest



Seoul Bike Data Best Performances (2L)

Performance (Python)

- 2-Layer Neural Network:
  Best R^2: 0.6602
  Activation and Optimizer: Activation: ReLU, Optimizer: SGD
  In-Sample MSE: 151686.234375
  In-Sample RMSE: 389.46917724609375
  In-Sample R^2: 0.6352
  Validation MSE: 149774.140625
  Validation RMSE: 387.00665283203125
  Validation R^2: 0.6405
  Cross-Validation MSE: 141125.546875
  Cross-Validation RMSE: 375.6151428222656
  Cross-Validation R^2: 0.6602

Performance (Scala)

- Tanh, eta: 0.001, epochs: 400
  - R-Squared: 0.692
  - Adjusted R-Squared: 0.691
  - MSE: 127995
  - RMSE: 357.763
  - MAE: 247.955

Seoul Bike Data Best Performances (3L)

Performance (Python)

- 3-Layer Neural Network:
  Best R^2: 0.8425
  Activation and Optimizer: Activation: SELU, Optimizer: SGD

  In-Sample MSE: 90175.5234375
  In-Sample RMSE: 300.2923889160156
  In-Sample R^2: 0.7831
  Validation MSE: 94231.7421875
  Validation RMSE: 306.9718933105469
  Validation R^2: 0.7738
  Cross-Validation MSE: 65384.8125
  Cross-Validation RMSE: 255.5073699951172
  Cross-Validation R^2: 0.8425

Performance (Scala)

- eLU, id, eta: 0.001, epochs=400
  - R-Squared: 0.861
  - Adjusted R-Squared: 0.860
  - MSE: 59892.1
  - RMSE: 244.729
  - MAE: 168.341

Seoul Bike Data Best Performances (XL)

Performance (Python)

- XL Neural Network:
  Best R^2: 0.8753
  Activation and Optimizer: Activation: SELU, Optimizer: SGD
  In-Sample MSE: 74270.4140625
  In-Sample RMSE: 272.5260009765625
  In-Sample R^2: 0.8214
  Validation MSE: 79367.546875
  Validation RMSE: 281.72247314453125
  Validation R^2: 0.8095
  Cross-Validation MSE: 51764.8515625
  Cross-Validation RMSE: 227.0684814453125
  Cross-Validation R^2: 0.8753

Performance (Scala)

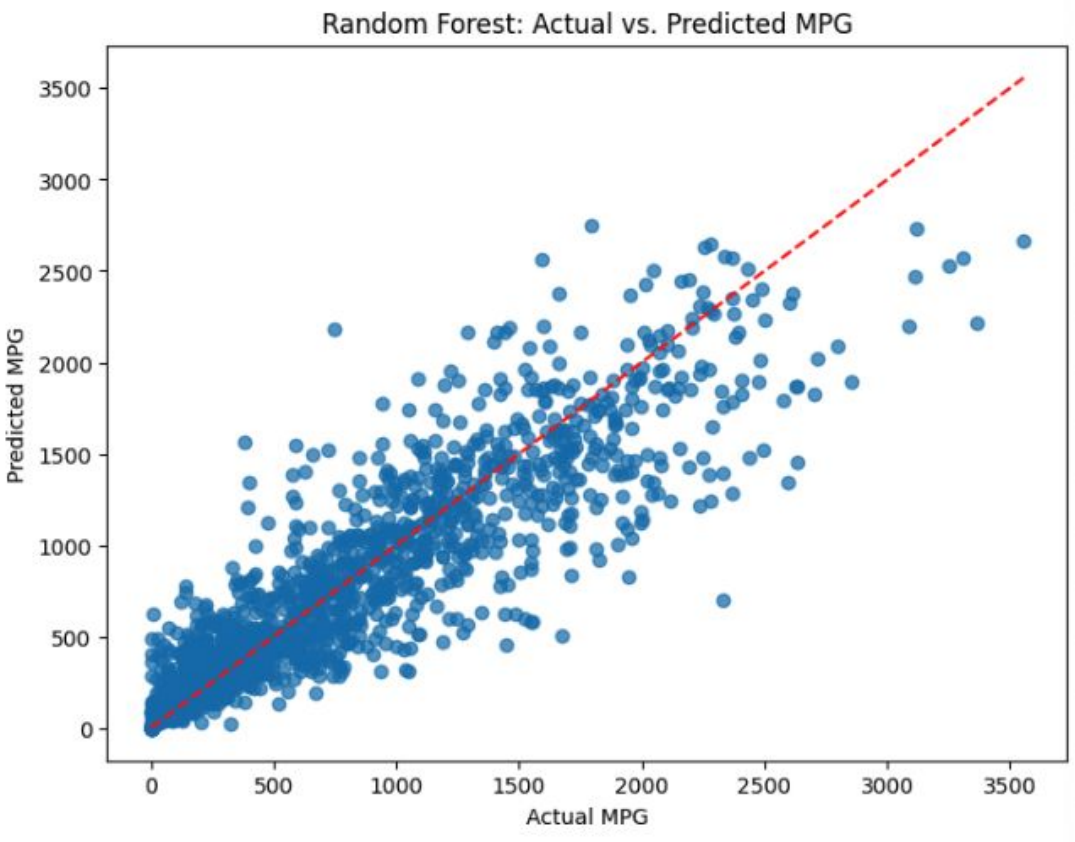
- reLU, reLU, id
  - R-Squared: 0.867
  - Adjusted R-Squared: 0.866
  - MSE: 55240.9
  - RMSE: 235.034
  - MAE: 159.035

Seoul Bike Data using Random forest

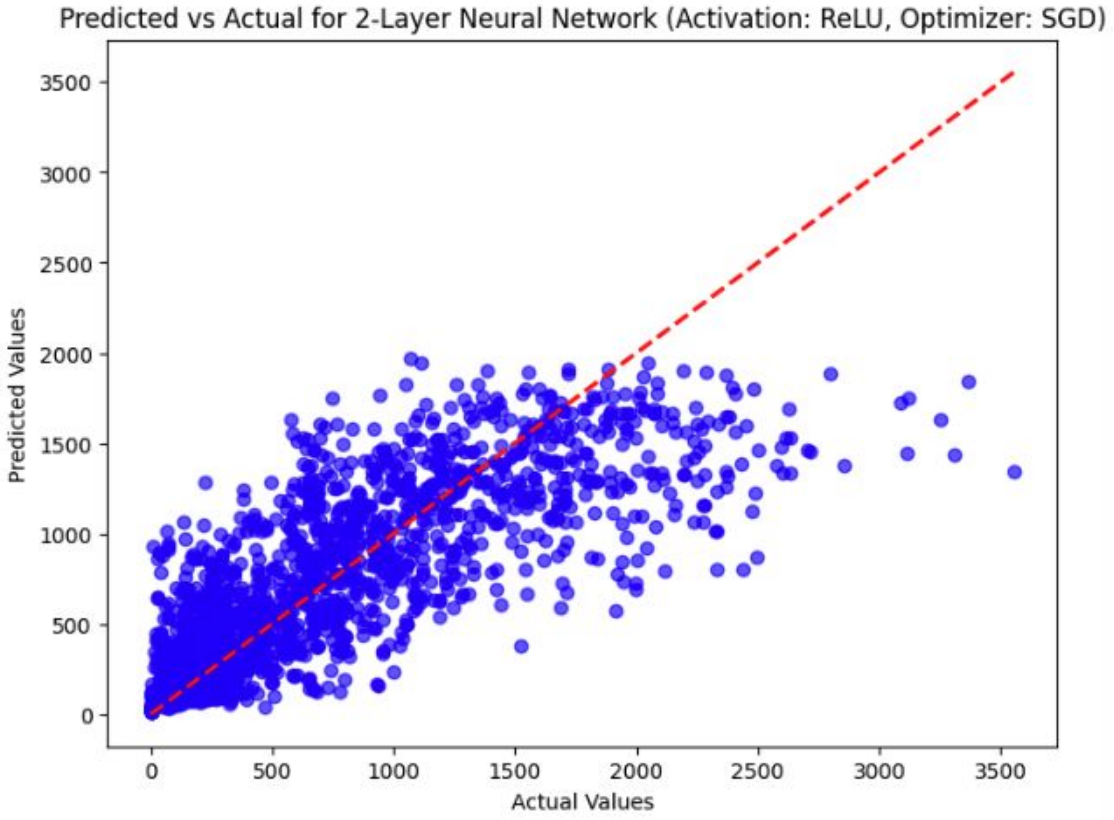
Random Forest Test MSE : 76185.7484

Random Forest Test RMSE: 276.0177

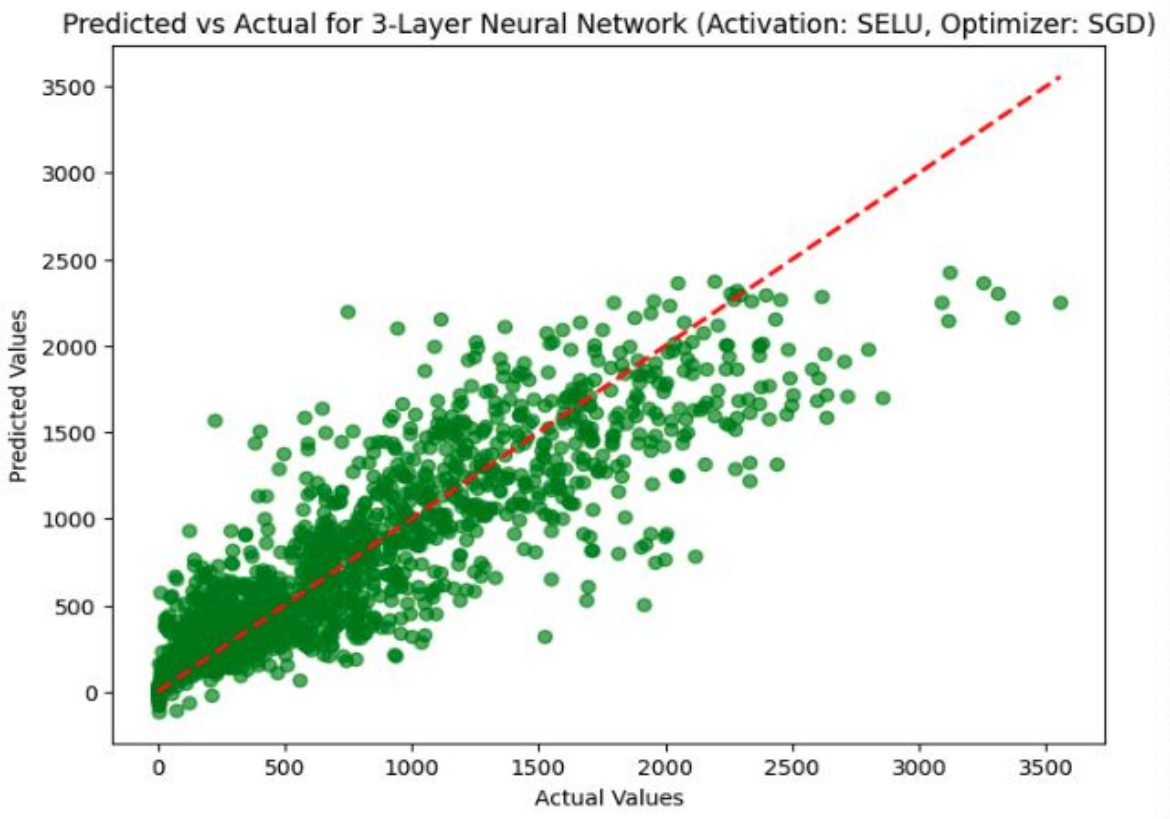
Random Forest Test R^2 : 0.8171



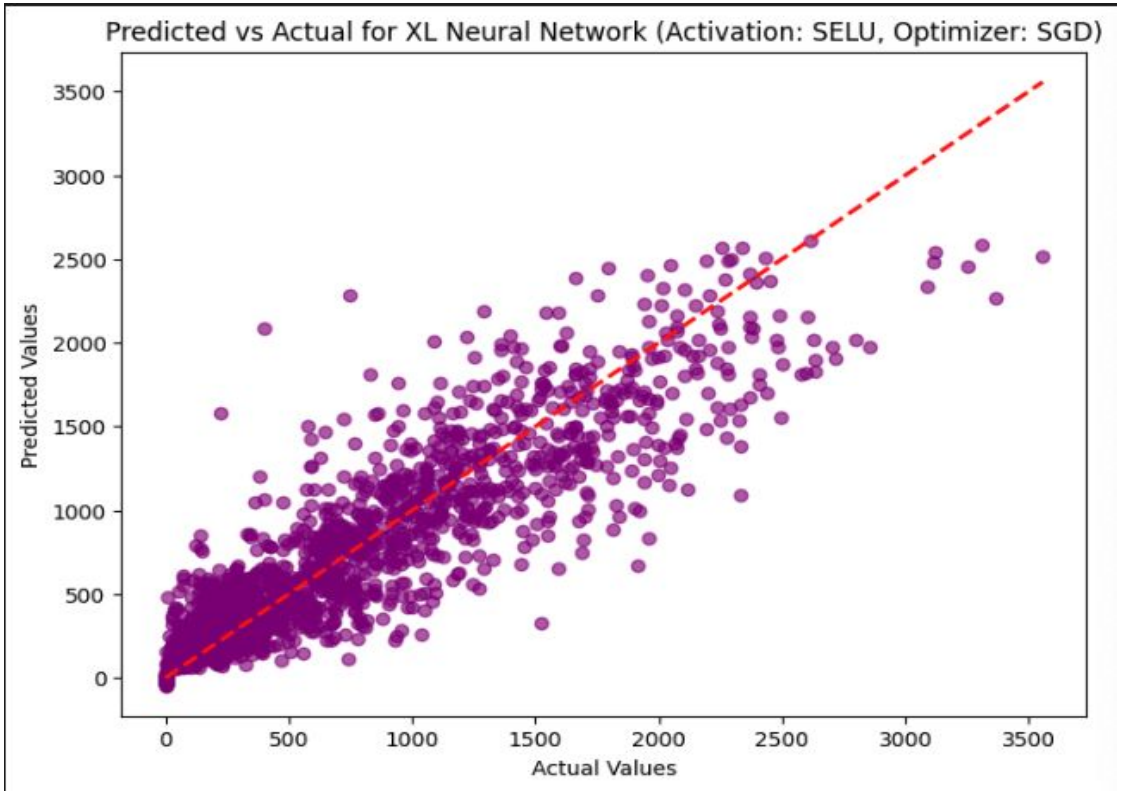
Seoul Bike Data Plots (Python, 2L)



Seoul Bike Data Plots (Python, 3L)

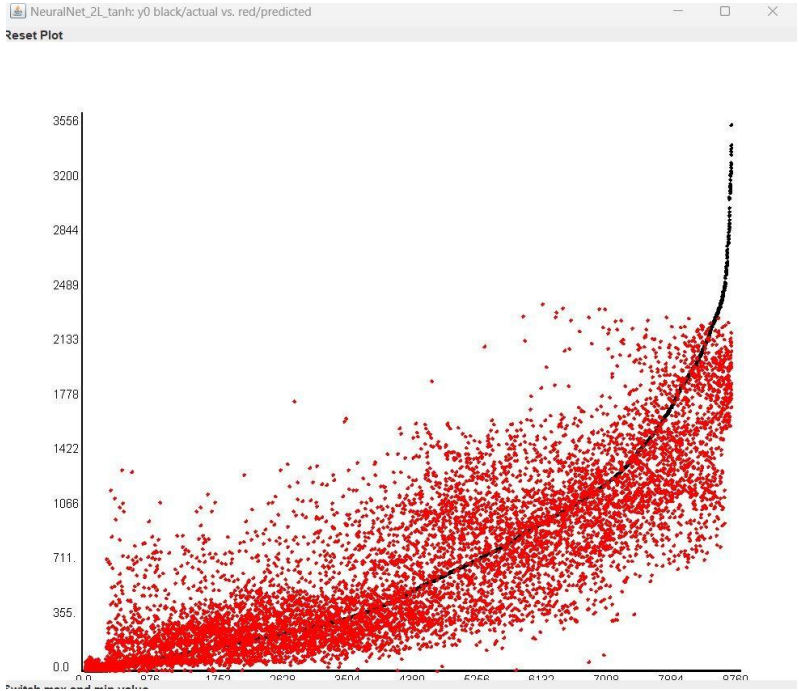


Seoul Bike Data Plots (Python, XL)

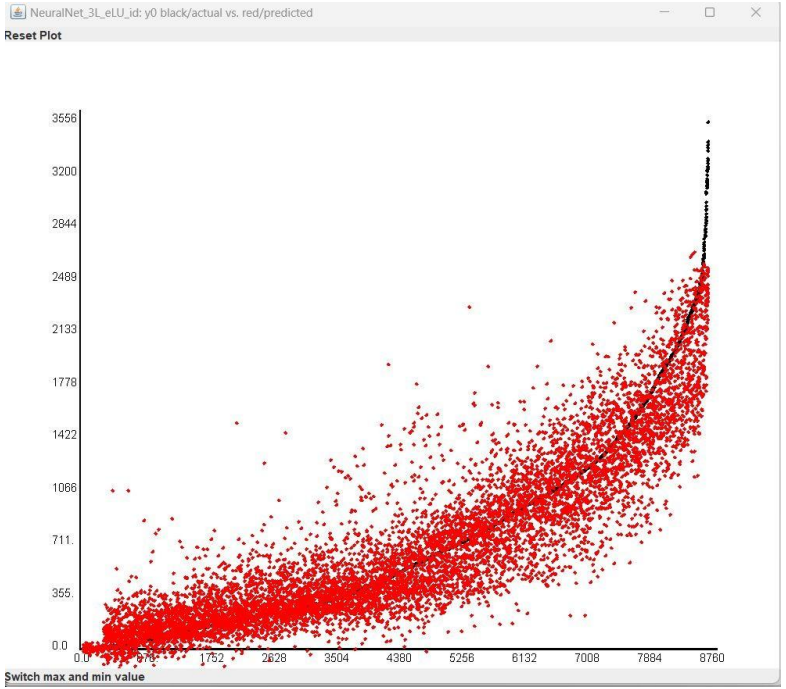


Seoul Bike Data Plots (Scala)

Neural Net 2L

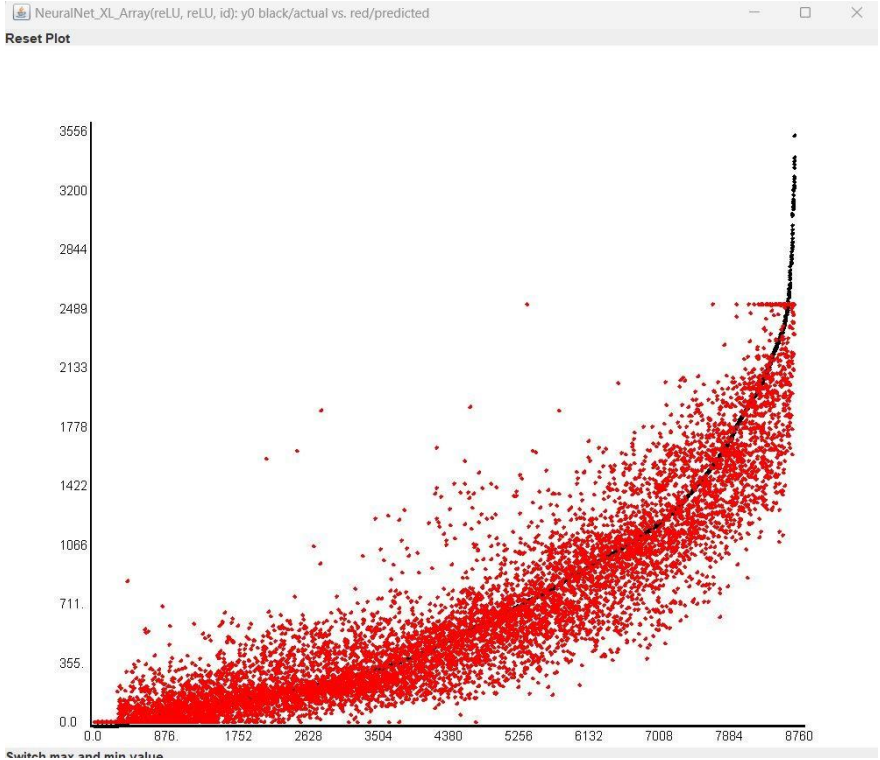


Neural Net 3L

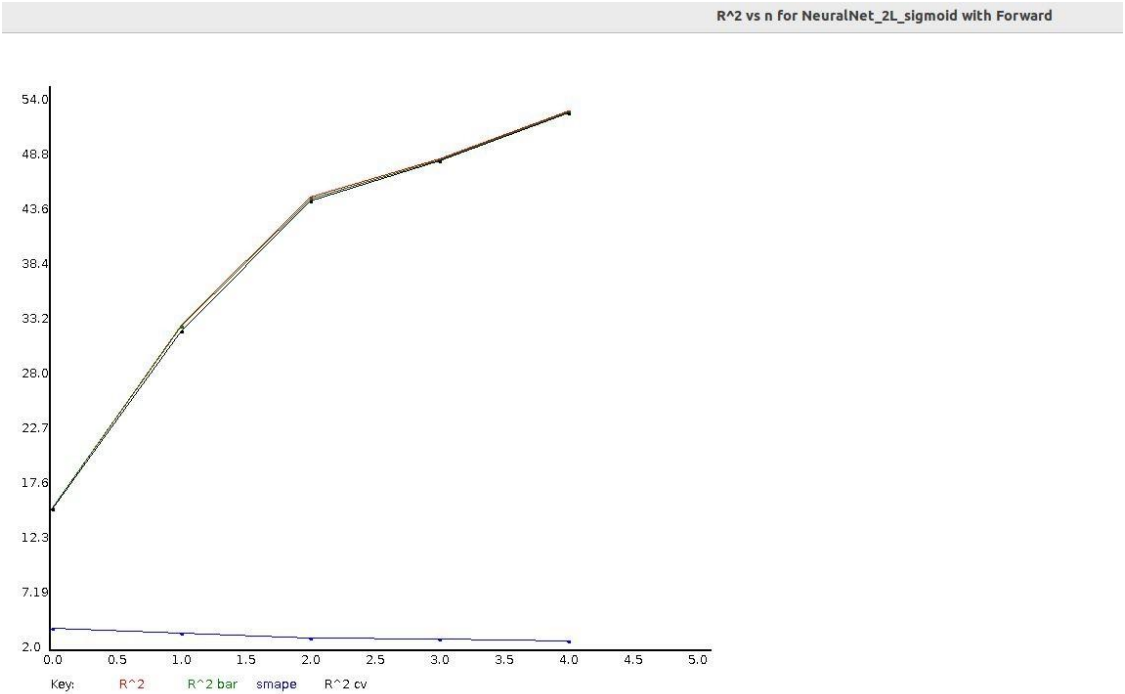


Seoul Bike Data Plots (Scala)

Neural Net XL



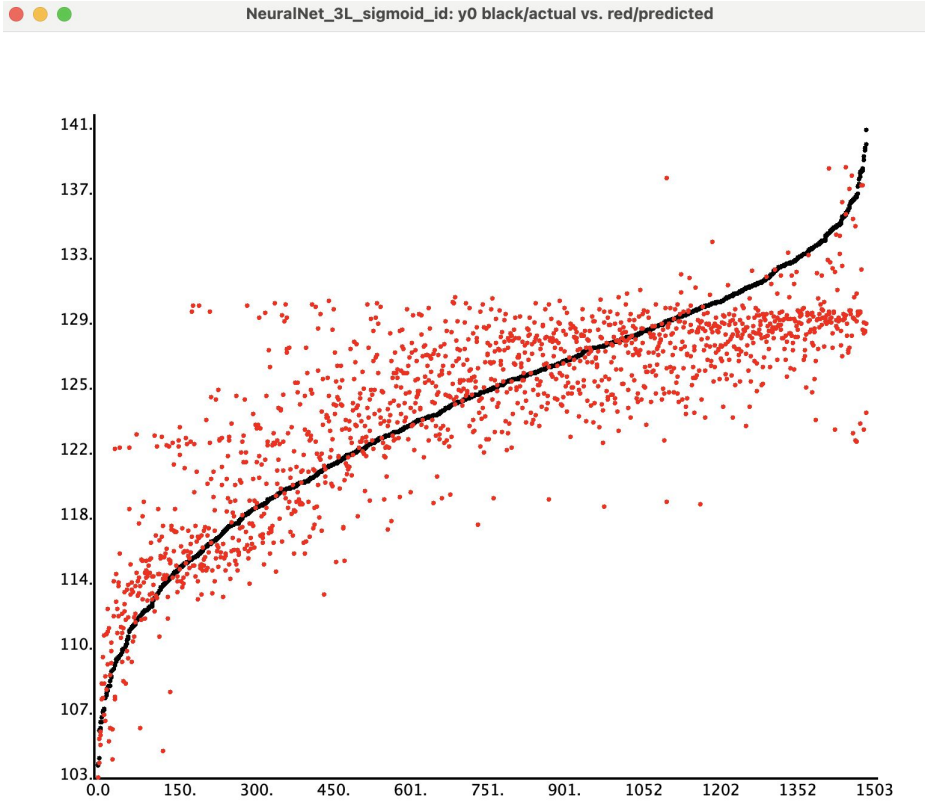
Airfoil 2L



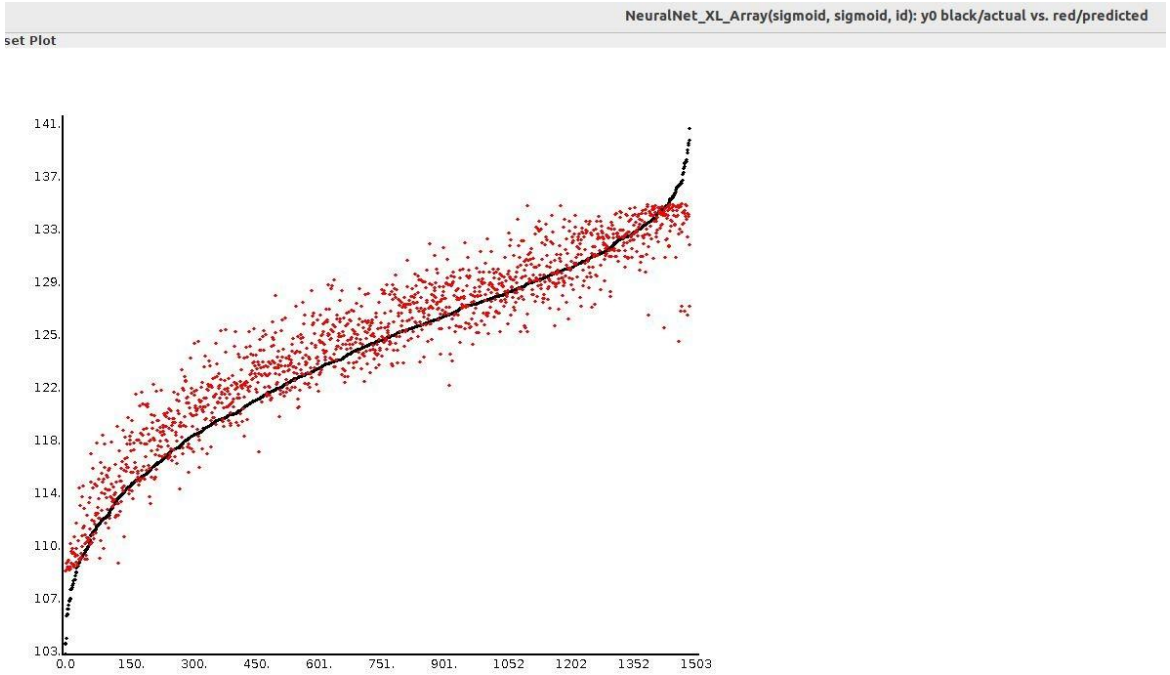
Sigmoid 2L:
R2- 0.533
Adj R2- 0.531

Airfoil 3L

Sigmoid
rSq -> 0.745438
rSqBar -> 0.744589

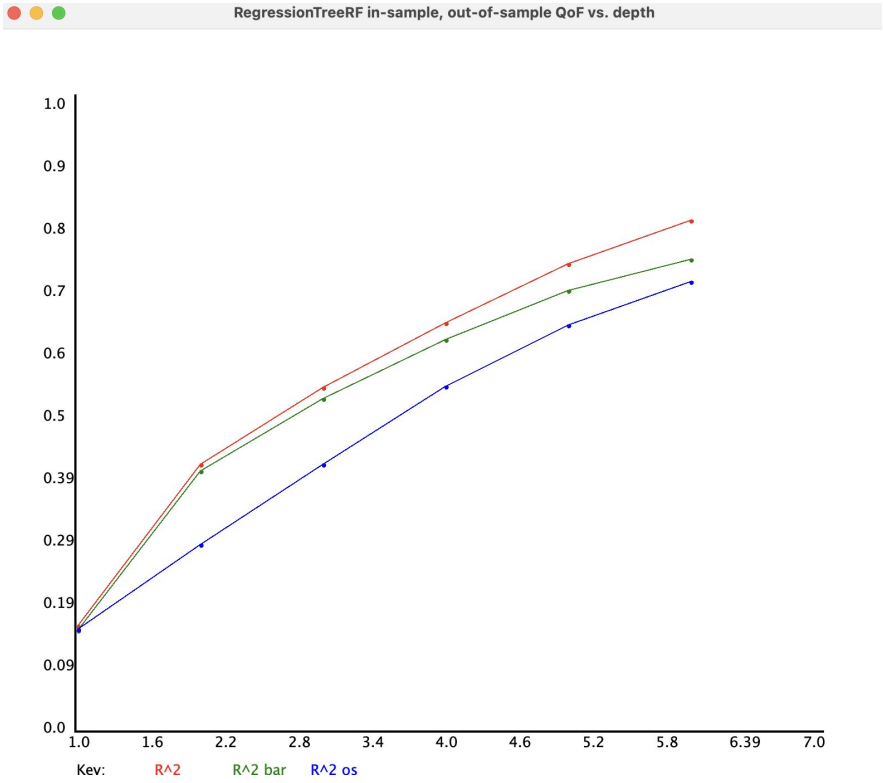


Airfoil



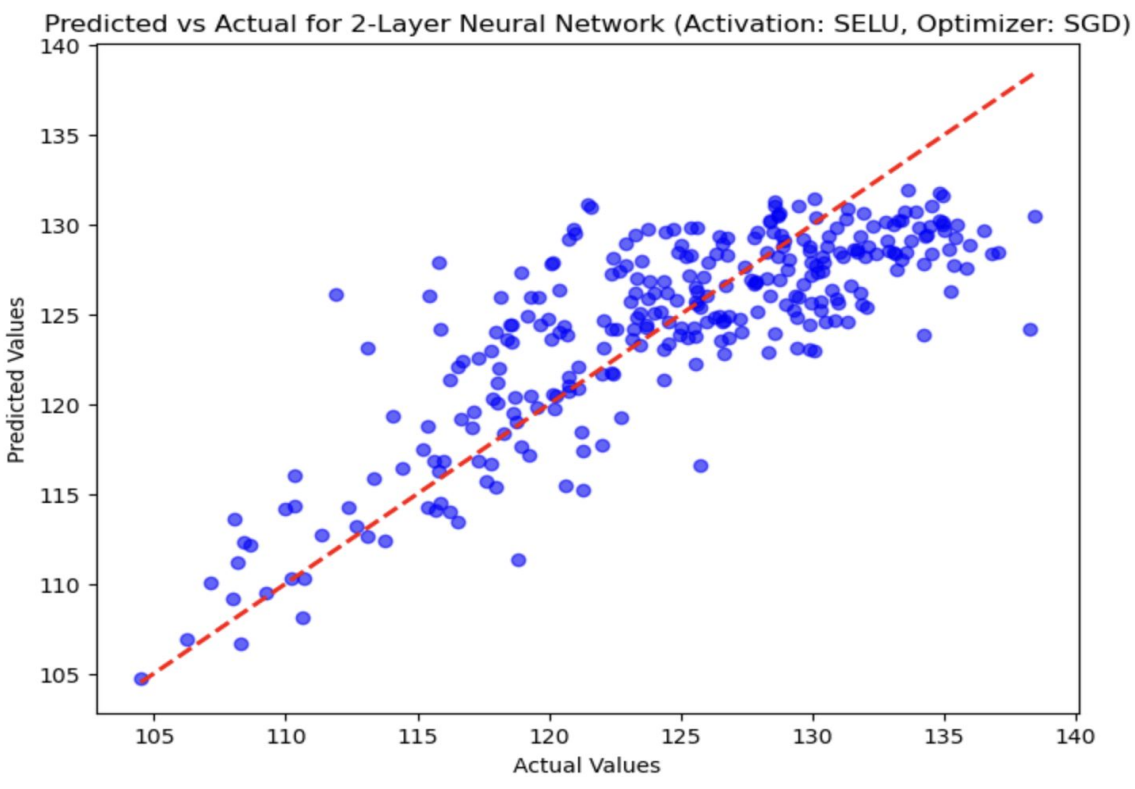
Sigmoid and ID XL:
R2- 0.922
Adj R2- 0.922

Airfoil



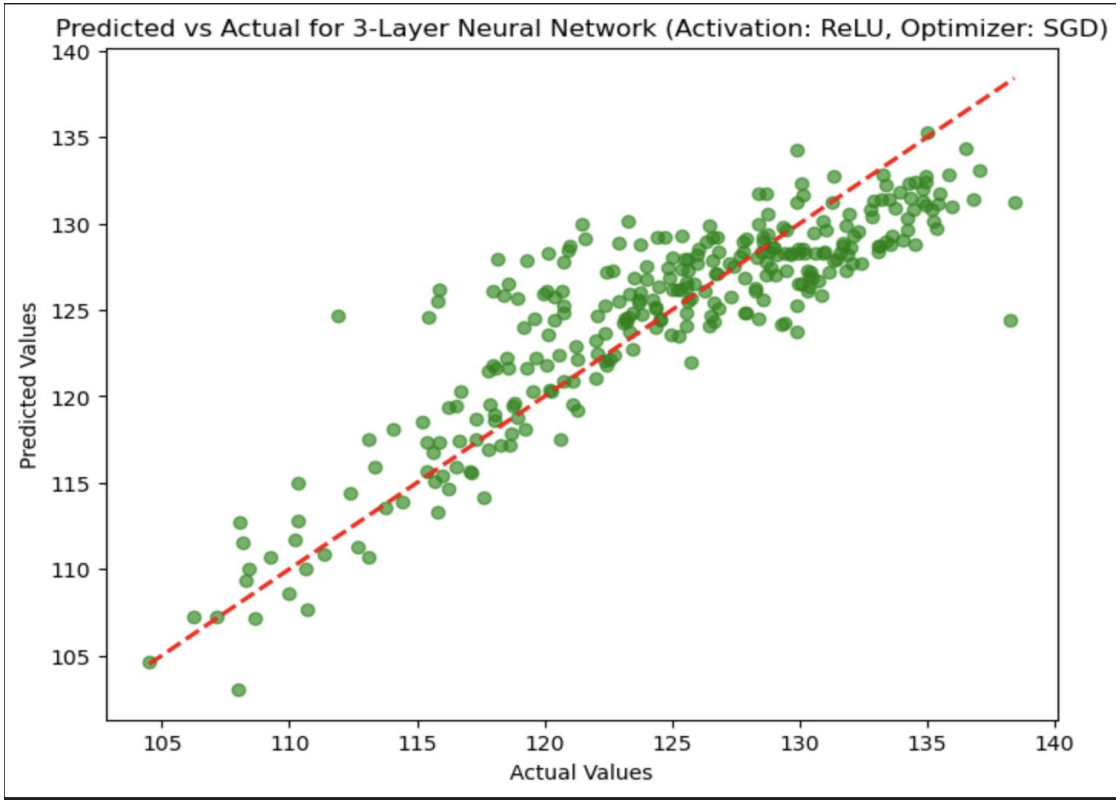
RF regressor
R2- 0.820

Airfoil



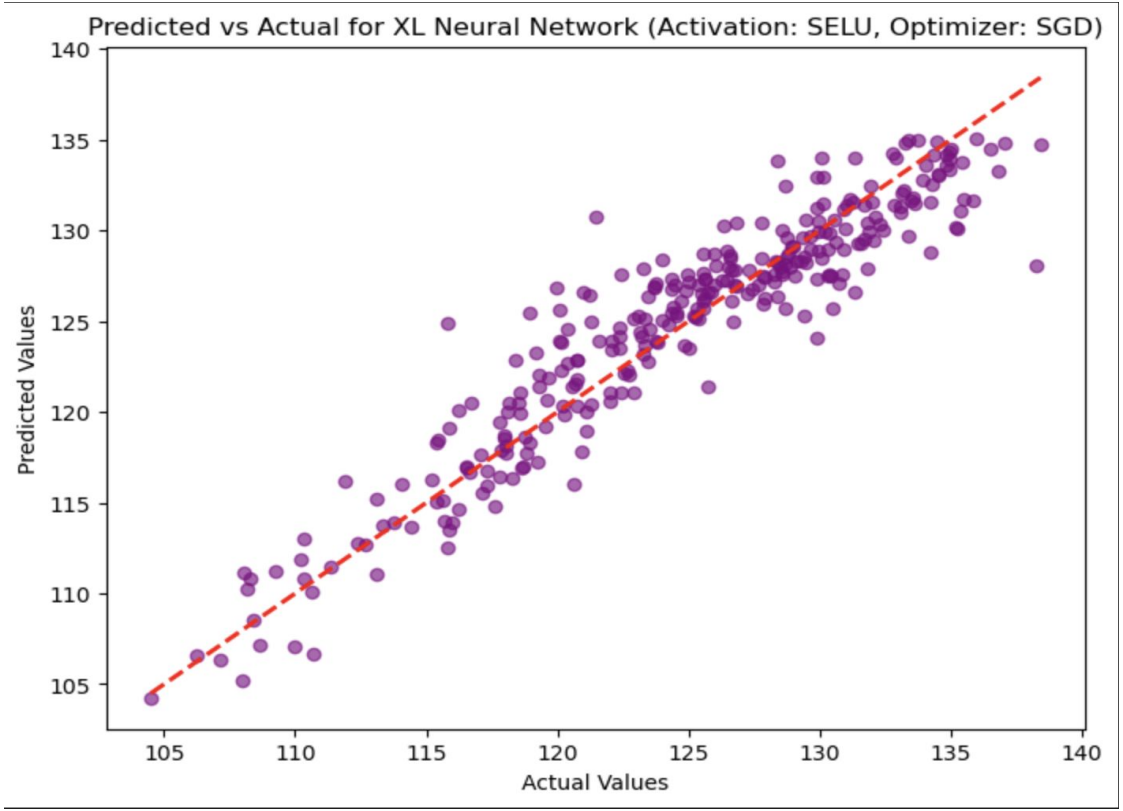
Python:
2L:
SELU,SGD
R2:0.6960.

AirFoil



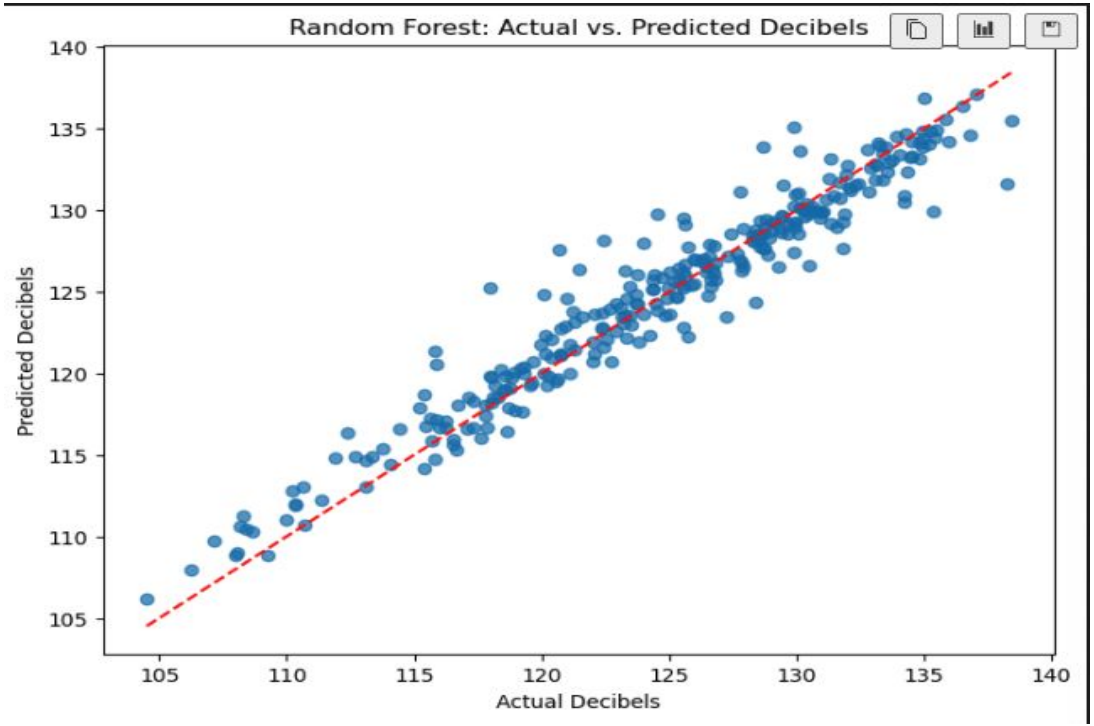
Python:
3L:
ReLU,SGD
R2:0.8630.

Airfoil



Python:
XL:
SELU,SGD
R2:0.9323.

Airfoil



Python:
Random Forest Regressor:

R2:0.9225.

Thank You