Summary autompg

Methodology

* The dataset itself is imported from the Scalation package
* Both in Scala and Python, mean squared error (MSE) was used as the loss function.
* In Python, the package of Keras used to build and run the Neural Nets.
* In Python, the Adam optimizer was used. In Scala, the Optimizer Object was used. Both languages the batch size was set to 32.
* In Python a preliminary run of each Neural Net was done with epochs set at 500 to find the most optimal epoch range for a given activation function and learning rate. This was determined by finding the lowest MSE value at a given epoch. An example is shown below.
* In Scala, the epochs were set to 200 as the findings from the Python version of the code showed 200 obtained a reasonable minimum for each neural net.
* In Python, the number of units in each layer was set to 24 while in Scala, Scalation used a formula to determine the number of units in each layer.
* In both languages, the activation functions of sigmoid, reLU and eLU were used at 0.001,0.01 and 0.1 learning rates.
* In both languages, a comparison was made between performing Forward Feature selection on the model prior to training and testing the Neural Net or directly feeding the data set into the Neural Net.

Feature Indexes

* 0- Cylinders
* 1 - Displacement
* 2 - Horsepower
* 3 - Weight
* 4 - Acceleration
* 5 - Model Year
* 6 - Origin
* 7 - mpg

**Table 1.** 2 Layer Neural Net with Forward Feature Selection in Scala

| Activation Fn | Learning Rate | Feature Index Selected | R2 Bar | R2 CV |
| --- | --- | --- | --- | --- |
| sigm | 0.1 | 0, 1, 6, 4, 3, 2, 5 | .793461 | 79.3461 |
| sigm | 0.01 | 0, 1, 2, 6, 4, 3, 5 | .463916 | 68.237 |
| sigm | 0.001 | 0, 6, 1, 5, 3, 4, 2 | -.890071 | -1.56678 |
| relu | 0.1 | 0, 4, 6, 5, 2, 1, 3 | .806167 | 79.7562 |
| relu | 0.01 | 0, 4, 6, 5, 1, 2, 3 | .804846 | 79.9259 |
| relu | 0.001 | 0, 4, 6, 5, 2, 3, 1 | .727379 | 79.221 |
| elu | 0.1 | 0, 4, 6, 5, 2, 1, 3 | .805868 | 80.0875 |
| elu | 0.01 | 0, 4, 6, 5, 1, 2, 3 | .803795 | 79.9239 |
| elu | 0.001 | 0, 4, 6, 5, 1, 3, 2 | -.376873 | 75.1549 |

**Table 2.** 2 Layer Neural Net in Scala

| Activation Fn | Learning Rate | CV R2 | R2 Bar | MSE |
| --- | --- | --- | --- | --- |
| sigm | 0.1 | 0.803058 | 0.799989 | 0.00844971 |
| sigm | 0.01 | 0.339961 | 0.329675 | 0.0283187 |
| sigm | 0.001 | -1.49031 | -1.52912 | 0.106846 |
| relu | 0.1 | 0.823212 | 0.820457 | 10.7234 |
| relu | 0.01 | 0.819965 | 0.81716 | 10.9203 |
| relu | 0.001 | 0.583183 | 0.576687 | 25.2828 |
| elu | 0.1 | 0.82318 | 0.820425 | 10.7253 |
| elu | 0.01 | 0.817109 | 0.814259 | 11.0936 |
| sigm | 0.1 | 0.803058 | 0.799989 | 0.00844971 |

**Table 3**. 3 Layer Neural Net with Forward Selection in Scala

| Activation Fn | Learning Rate | Feature Index Selected | CV R2 | R2 Bar |
| --- | --- | --- | --- | --- |
| igm | 0.1 | 0, 4, 6, 3, 5, 2, 1 | .874931 | 88.0455 |
| sigm | 0.01 | 0, 4, 6, 3, 5, 2, 2 | .863329 | 86.5124 |
| sigm | 0.001 | 0, 6, 1, 4, 2, 3, 5 | .736761 | 81.546 |
| relu | 0.1 | 0, 1, 6, 4, 3, 2, 5 | .85131 | 85.3965 |
| relu | 0.01 | 0, 4, 6, 5, 3, 2, 1 | .813336 | 80.6446 |
| relu | 0.001 | 0, 4, 6, 5, 3, 2, 1 | .816882 | 81.3303 |
| elu | 0.1 | 0, 4, 6, 3, 5, 2, 1 | .90316 | 87.9927 |
| elu | 0.01 | 0, 4, 6, 5, 1, 3, 2 | .87014 | 87.3269 |
| elu | 0.001 | 0, 4, 6, 3, 1, 2, 5 | .863421 | 85.6921 |

**Table 4**. 3 Layer Neural Net in Scala

| Activation Fn | Learning Rate | CV R2 | R2 Bar | MSE |
| --- | --- | --- | --- | --- |
| sigm | 0.1 | 0.866101 | 0.863667 | 8.12188 |
| sigm | 0.01 | 0.850965 | 0.848255 | 9.04002 |
| sigm | 0.001 | 0.87068 | 0.868334 | 0.868334 |
| relu | 0.1 | 0.825767 | 0.822599 | 10.5684 |
| relu | 0.01 | 0.826957 | 0.823811 | 10.4962 |
| relu | 0.001 | 0.827008 | 0.823863 | 10.4931 |
| elu | 0.1 | 0.880012 | 0.87783 | 7.27811 |
| elu | 0.01 | 0.867558 | 0.86515 | 8.0335 |
| elu | 0.001 | (0.871047 | 0.868703 | 7.82187 |

**Table 5**. 4 Layer Neural Net with Forward Selection in Scala

| Activation Fn | Learning Rate | Feature Index Selected | CV R2 | R2 Bar |
| --- | --- | --- | --- | --- |
| sigm | 0.1 | 0, 4, 6, 3, 5, 2, 1 | .17718 | 70.4817 |
| sigm | 0.01 | 0, 4, 6, 3, 5, 2, 2 | .0528021 | 8.47886 |
| sigm | 0.001 | 0, 6, 1, 4, 2, 3, 5 | -277.17 | -6.2453 |
| relu | 0.1 | 0, 1, 6, 4, 3, 2, 5 | .852273 | 82.2187 |
| relu | 0.01 | 0, 4, 6, 5, 3, 2, 1 | .811519 | 80.3514 |
| relu | 0.001 | 0, 4, 6, 5, 3, 2, 1 | .817523 | 81.5438 |
| elu | 0.1 | 0, 4, 6, 3, 5, 2, 1 | .887861 | 88.4036 |
| elu | 0.01 | 0, 4, 6, 5, 1, 3, 2 | .866947 | 87.3002 |
| elu | 0.001 | 0, 4, 6, 3, 1, 2, 5 | .863352 | 85.8119 |

**Table 6**. 4 Layer Neural Net in Scala

| Activation Fn | Learning Rate | CV R2 | R2 Bar | MSE |
| --- | --- | --- | --- | --- |
| sigm | 0.1 | 0.154467 | 0.139093 | 0.0362773 |
| sigm | 0.01 | 0.066944 | 0.0499793 | 0.0400324 |
| sigm | 0.001 | -4.1943 | -4.2888 | 0.222863 |
| relu | 0.1 | 0.847684 | 0.844915 | 9.23898 |
| relu | 0.01 | 0.826041 | 0.822878 | 10.5518 |
| relu | 0.001 | 0.827709 | 0.824577 | 10.4506 |
| elu | 0.1 | 0.872917 | 0.870607 | 7.70843 |
| elu | 0.01 | 0.867382 | 0.864971 | 8.0442 |
| elu | 0.001 | 0.871087 | 0.868743 | 7.81945 |

**Table 7**. 2 Layer Neural Net with Forward Selection in Python

| Activation Fn | Learning Rate | Feature Index Selection | CV R2 | R2 Bar |
| --- | --- | --- | --- | --- |
| sigm | 0.1 | 5,7,8 | -0.139675 | 0.349906 |
| sigm | 0.01 | 5,7,8 | -0.410933 | 0.052919 |
| sigm | 0.001 | 5,7,8 | -4.594451 | 0.010148 |
| relu | 0.1 | 5,7,8 | -0.147744 | 0.34214 |
| relu | 0.01 | 5,7,8 | -0.367637 | 0.014484 |
| relu | 0.001 | 5,7,8 | -4.591899 | 0.017171 |
| elu | 0.1 | 5,7,8 | -0.154673 | 0.314381 |
| elu | 0.01 | 5,7,8 | -0.428767 | 0.094859 |
| elu | 0.001 | 5,7,8 | -4.449582 | 0.014532 |

**Table 8**. 2 Layer Neural Net in Python

| Activation Fn | Learning Rate | R2 | CV R2 | R2 Bar |
| --- | --- | --- | --- | --- |
| sigm | 0.1 | 5,7,8 | -0.488543 | -0.006601 |
| sigm | 0.01 | 5,7,8 | -0.409423 | -0.006601 |
| sigm | 0.001 | 5,7,8 | -0.364195 | -0.006601 |
| relu | 0.1 | 5,7,8 | 0.610617 | 0.26149 |
| relu | 0.01 | 5,7,8 | 0.611785 | 0.196392 |
| relu | 0.001 | 5,7,8 | 0.526951 | 0.270002 |
| elu | 0.1 | 5,7,8 | 0.481787 | 0.211575 |
| elu | 0.01 | 5,7,8 | 0.035867 | 0.187672 |
| elu | 0.001 | 5,7,8 | 0.639334 | 0.274752 |

**Table 9**. 3 Layer Neural Net with Forward Selection in Python

| Activation Fn | Learning Rate | Feature Index Selection | CV R2 | R2 Bar |
| --- | --- | --- | --- | --- |
| sigm | 0.1 | 5,7,8 | -0.461189 | -0.006601 |
| sigm | 0.01 | 5,7,8 | -0.429653 | -0.006601 |
| sigm | 0.001 | 5,7,8 | -0.422419 | -0.006601 |
| reLu | 0.1 | 5,7,8 | -0.170269 | -0.006601 |
| reLu | 0.01 | 5,7,8 | 0.619104 | 0.255028 |
| reLu | 0.001 | 5,7,8 | 0.594927 | 0.274125 |
| elu | 0.1 | 5,7,8 | -1.513295 | 0.11644 |
| elu | 0.01 | 5,7,8 | 0.297046 | 0.210559 |
| elu | 0.001 | 5,7,8 | 0.272187 | 0.283957 |

**Table 10.** 3 Layer Neural Net in Python

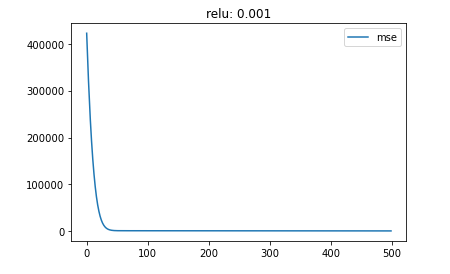
| Activation Fn | Learning Rate | R2 | CV R2 | R2 Bar |
| --- | --- | --- | --- | --- |
| sigm | 0.1 | 0.00E+00 | -0.523985 | -0.020067 |
| sigm | 0.01 | 0.00E+00 | -0.403388 | -0.020067 |
| sigm | 0.001 | 2.98E-08 | -0.41504 | -0.020067 |
| relu | 0.1 | 5.88E-02 | -0.09143 | 0.039894 |
| relu | 0.01 | 2.37E-01 | 0.500093 | 0.222101 |
| relu | 0.001 | 2.93E-01 | 0.500941 | 0.279167 |
| elu | 0.1 | 2.50E-01 | 0.4107 | 0.2352 |
| elu | 0.01 | 2.89E-01 | 0.542141 | 0.274322 |
| elu | 0.001 | 2.16E-01 | 0.248639 | 0.200365 |

**Table 11**. 4 Layer Neural Net with Forward Selection in Python

| Activation Fn | Learning Rate | Feature Index Selection | CV R2 | R2 Bar |
| --- | --- | --- | --- | --- |
| sigm | 0.1 | 5,7,8 | -0.461189 | -0.006601 |
| sigm | 0.01 | 5,7,8 | -0.429653 | -0.006601 |
| sigm | 0.001 | 5,7,8 | -0.422419 | -0.006601 |
| relu | 0.1 | 5,7,8 | -0.170269 | -0.006601 |
| relu | 0.01 | 5,7,8 | 0.619104 | 0.255028 |
| relu | 0.001 | 5,7,8 | 0.594927 | 0.274125 |
| elu | 0.1 | 5,7,8 | -1.513295 | 0.11644 |
| elu | 0.01 | 5,7,8 | 0.297046 | 0.210559 |
| elu | 0.001 | 5,7,8 | 0.272187 | 0.283957 |

**Table 12**. 4 Layer Neural Net in Python

| Activation Fn | Learning Rate | R2 | CV R2 | R2 Bar |
| --- | --- | --- | --- | --- |
| sigm | 0.1 | -1.49E-08 | -0.516527 | -0.020067 |
| sigm | 0.01 | 0.00E+00 | -0.418862 | -0.020067 |
| sigm | 0.001 | 0.00E+00 | -0.419258 | -0.020067 |
| relu | 0.1 | 0.00E+00 | -0.009573 | -0.020067 |
| relu | 0.01 | 2.46E-01 | 0.297114 | 0.231054 |
| relu | 0.001 | 2.80E-01 | 0.437474 | 0.265593 |
| elu | 0.1 | 0.00E+00 | -1.010899 | -0.020067 |
| elu | 0.01 | 2.50E-01 | -0.064635 | 0.234748 |
| elu | 0.001 | 2.87E-01 | 0.504221 | 0.272838 |



**Figure 1**: An example of seeing how many epochs it takes to minimize the MSE (the loss function used in each model). With the default number of epochs set at 500, this example shows that for the reLU activation function and 0.001 learning rate, the around 200 epochs is where the minimization in loss becomes stagnant.