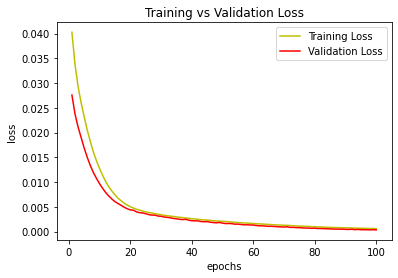
Mitchell Covey

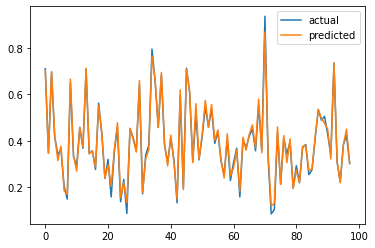
Final Report: CISC 489

We are trying to create a regression model to predict Boston housing prices with the housing.csv dataset. The implementation used experimented with different factors to see what would improve results. The first network is a base network with one layer. The second network adds another layer. The third adds dropouts. The fourth changes activation functions. The fifth network takes all those into account and determines the best network.

Network Structure 1: One layer with 100 Neurons. Relu is our activation function.

|  |  |  |
| --- | --- | --- |
| Optimizers | MSE | RMSE |
| SGD | 2.2097108517777553 | 0.15016014756285884 |
| adam | 0.025753774365778415 | 0.016210910399398634 |
| rmsprop | 0.052321439979940945 | 0.023106108384868165 |
| Adadelta | 3.5233652844744068 | 0.18961199061354933 |
| Adagrad | 1.9442520999045303 | 0.14085207539075512 |

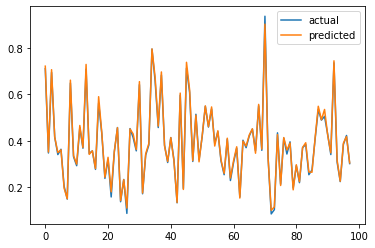
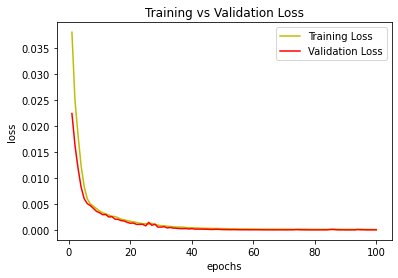




Adam was by far the best optimizer for this network. For our next model, we’ll try adding another layer.

Network Structure 2: Two layers with 100 neurons each. Relu is our activation function.

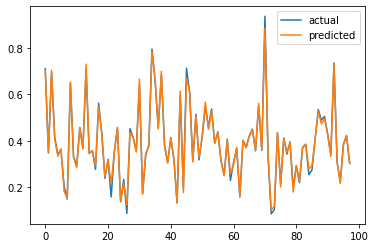
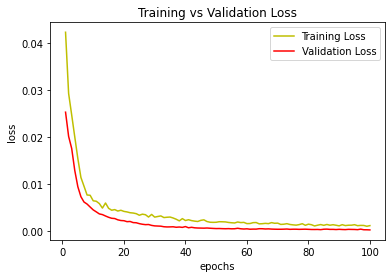
|  |  |  |
| --- | --- | --- |
| Optimizers | MSE | RMSE |
| SGD | 1.8580860813900433 | 0.1376955411587479 |
| adam | 0.007707782917853169 | 0.008868531225662035 |
| rmsprop | 0.03155135015939889 | 0.017943036327414806 |
| Adadelta | 2.896092201219713 | 0.17190683883119903 |
| Adagrad | 2.59085864560888 | 0.16259561223617924 |



Adam once again dominated every other optimizer. To switch things up, we’ll add some dropouts.

Network Structure 3: Two Layers with 100 neurons each. Dropouts added after each layer. Relu is still our activation function.

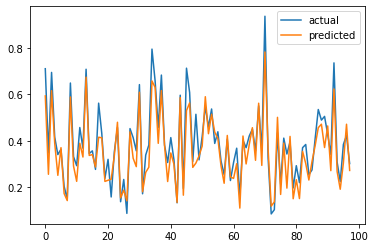
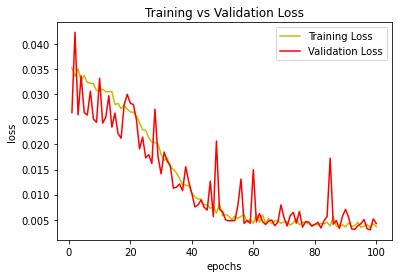
|  |  |  |
| --- | --- | --- |
| Optimizers | MSE | RMSE |
| SGD | 2.0257859333776977 | 0.1437751196642392 |
| adam | 0.019426030967267107 | 0.014079233139217487 |
| rmsprop | 0.07010552023965395 | 0.02674626051292106 |
| Adadelta | 3.074758494235329 | 0.17713014050655157 |
| Adagrad | 2.8520254557002267 | 0.17059396404436372 |



While adam still dominates, dropouts did not improve the model in the slightest. For our next model, we’re going to switch gears and try sigmoid as an activation function.

Network Structure 4: Two layers with 100 neurons each. Sigmoid is our activation function.

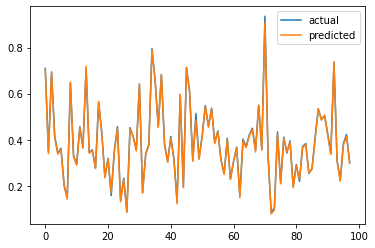
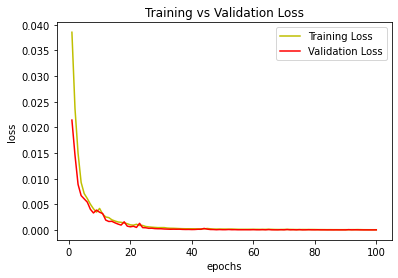
|  |  |  |
| --- | --- | --- |
| Optimizers | MSE | RMSE |
| SGD | 2.6246595584032084 | 0.16365280441186908 |
| adam | 0.25461676710585635 | 0.050971857692165494 |
| rmsprop | 0.2512288210053772 | 0.05063160473473118 |
| Adadelta | 4.1858513452834805 | 0.20667067723657187 |
| Adagrad | 2.614444723961012 | 0.16333403621828876 |



Rmsprop finally dethrones adam as the better optimizer. Relu with adam however gave us much better results, so the next model will lean into what worked and increase our layer count with relu.

Network Structure 5: Three dense layers, 200, 100, 50 neurons each. Relu is our activation function.

|  |  |  |
| --- | --- | --- |
| Optimizers | MSE | RMSE |
| SGD | 1.904736585348252 | 0.13941336953676434 |
| adam | 0.003803374818150412 | 0.006229763007048005 |
| rmsprop | 0.09657936923282709 | 0.03139273431356226 |
| Adadelta | 3.6883040855101226 | 0.1939993710675194 |
| Adagrad | 2.8290706648542066 | 0.16990605642153075 |



Adam’s numbers in this model are an improvement from network 2, our previous best. If I was going to implement any of the models, I would pick network 5. It seems adam and relu are a very good combination for decreasing the RMSE value.