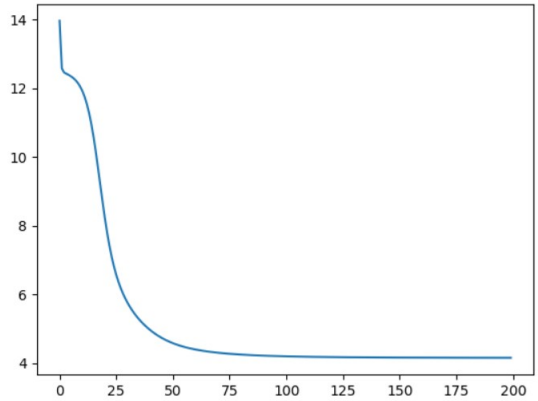
Dawid Sitnik, 2016118376

Q 2.

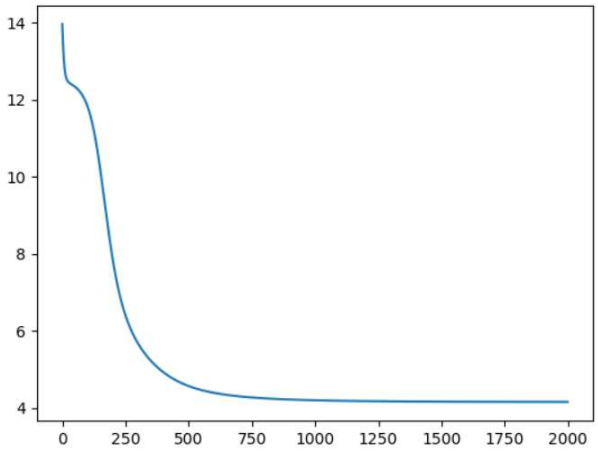
Neutral network: 2 input nodes, 1 hidden layer (4 nodes), 1 output node

Dataset: non-linear dataset of 100 data-points, created with Scikit library.

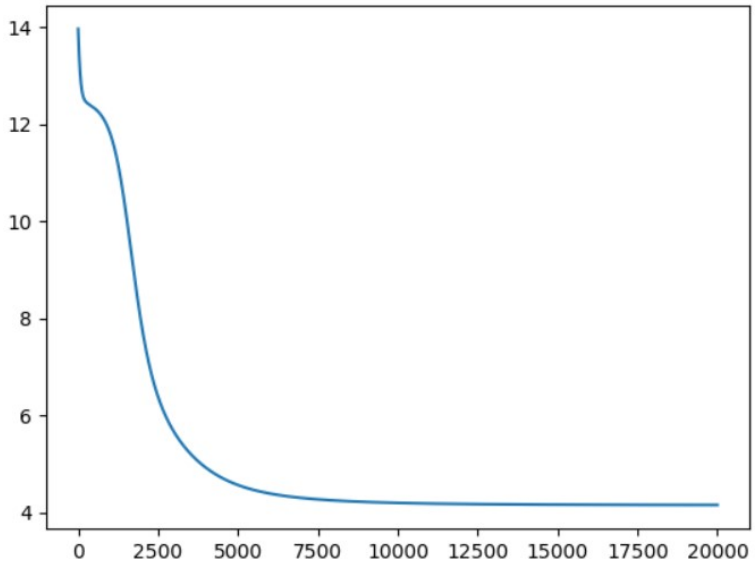
Sigmoid, learning rate = 0.1



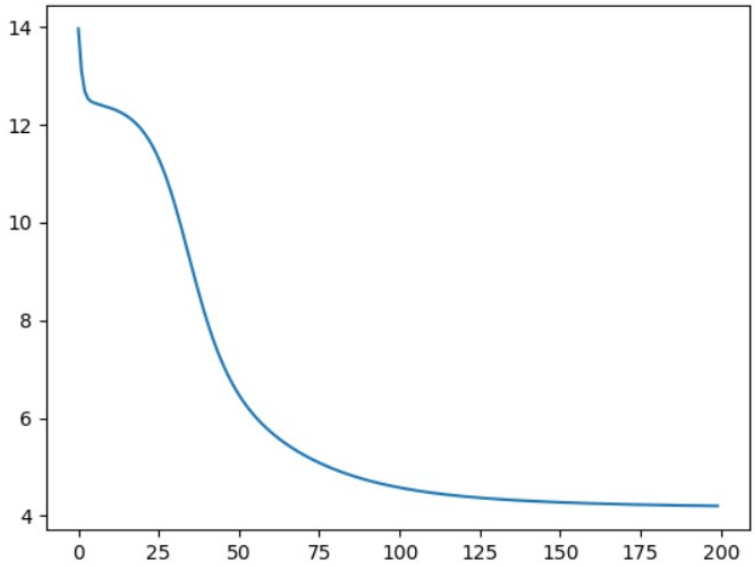
Sigmoid, learning rate = 0.01



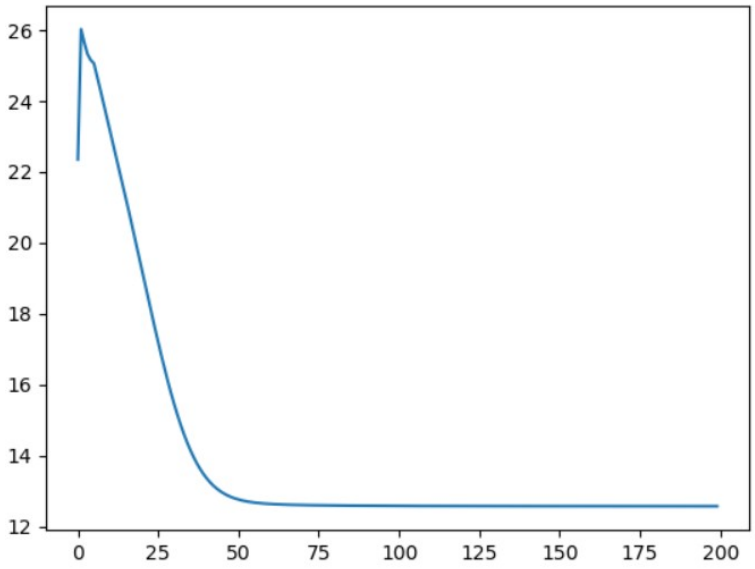
Sigmoid, learning rate = 0.001



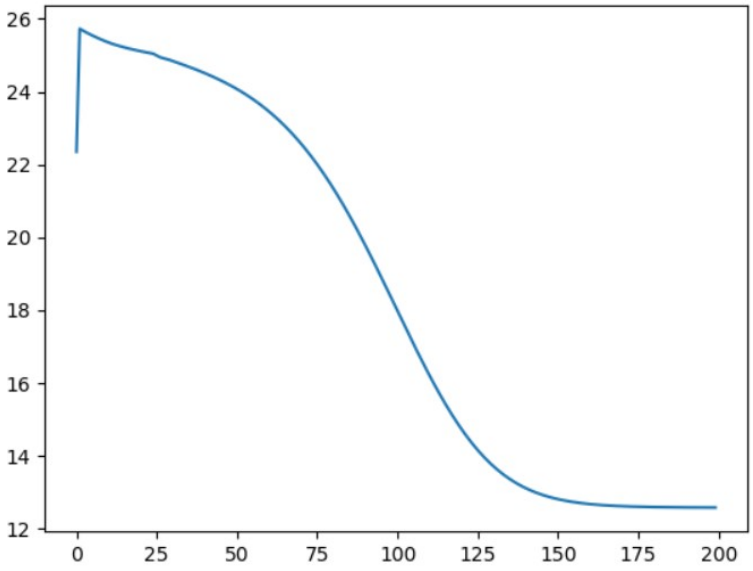
Sigmoid, learning rate = 0.05



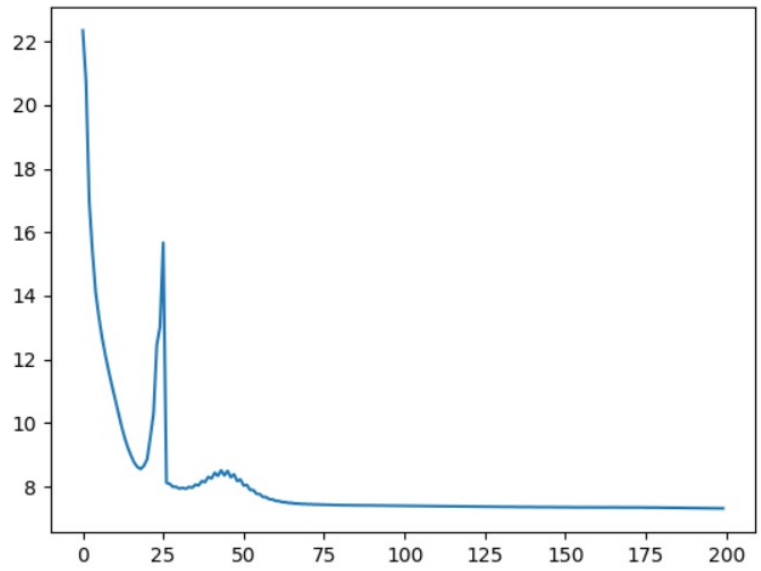
Leaky ReLU, learning rate = 0.1



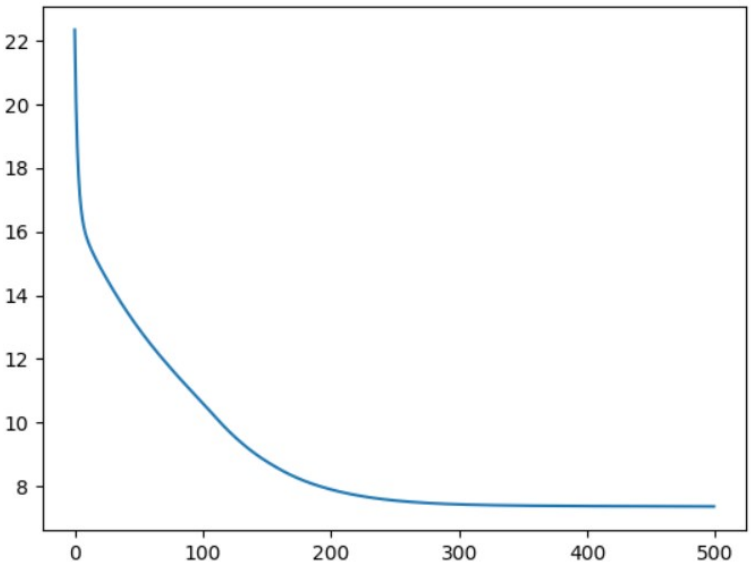
Leaky ReLU, learning rate = 0.05



Leaky ReLU, learning rate = 0.01



Leaky ReLU, learning rate = 0.001



Conclusion

As we can conclude from the plots, sigmoid as an activation function performs much better than leaky relu.

According to my dataset the best performence of signoid is acheived with learning rate of 0.1. In this case I obtained error rate of about 4 within about 100 iterations. Changing

learning rate to smaller it will also obtain error rate of 4, but it will take longer.

While using leaky relu results seem to have some errors (peaks on the plots). Whats more, the smallest possible rate is about 8 in the best case, which is still two times worse than signoid. The best result was obtained for learning rate of 0.01, while we got the smallest possible error rate after about 70 interations. Using bigger learning rates, the smallest possible error rate to obtain was about 12.