

Lab 2 Report

Zhenda Lu zhenda.lu@wisc.edu
Renjie Tang rtang37@wisc.edu

Objective

Use one hidden layer with ReLU and Sigmoid to predict the secondary protein. Use rectified Linear for the HU activation, and use sigmoid for the output units. After modifying our ANN many times with different method, we claim that 7 hidden units without any regulation can produce the best result. We can get an accuracy around 0.58, 0.59.

Hidden Units

We start with 3 hidden units, which gives us low accuracy. With using 10, 100, 1000, or other numbers of HUs, 7 HUs gives a stable and relatively better accuracy for test and tune result. In addition, the training speed explicitly decrease with the increase of the number of HUs; however, the increasing number of HUs does not guarantee higher accuracy or lower error rate.

Activation Function

As what included in the slides and piazza, we try to use ReLU or Sigmoid for the HU activation. With the number of HU as 5, the Sigmoid function does not perform as good as ReLU, and the accuracy with two Sigmoid is around 56%, which is not better than our original 58%. So we decided to stick with ReLU and only use Sigmoid for the output units.

Early Stopping

We monitor tune set and test set accuracy in many different situations. And we found out that stopping when both reach 58% could probably reduce overfitting.

Weight Decay

We've tried weight decay with lambda values: 10, 1, 0.1, 0.005. Weight decay greatly slows down the training process. And it gave back lower accuracy (40%). So we decided not to use it.

Momentum

We fixed beta value to be 0.9. Momentum regulation will cause the network to converge and overfit very fast. To offset the disadvantage, we tried to lower learning rate. It slowed the training down but didn't improve the test set accuracy. Our idea is that with momentum regulation, it allows us to test lower learning rate, which might give better local minima sometimes, but it is not obvious for this project.

Combine Momentum and Weight Decay

This gave back very low accuracy and bad training speed. The accuracy can be as low as about 20%. We suspect there exists a bug, but do not have time to prove and fix. We simply decided not to use them.

Conclusion and Further Thoughts

This ANN implies the strong connection between complexity of the design and training time. With more regulations or the number of HUs, it takes much longer time to get the result and compare different method. In the light of this, we believe that we better to understand and compare the mathematics behind different approach before deciding using any. The program serves its purpose to predict the secondary protein; however, it has many flaws. For one, some of the code are not general, which makes it hard to reuse. We would like to improve the code in the future.