Artificial Neural Network

Connor Wolfe and Jonathan Yau

October 2017

What is the area under ROC of the ANN with no hidden units on each dataset? Set the weight decay coefficient =0, and train to convergence. This approximates the perceptron, which uses a step function instead of a sigmoid. How does this compare to the decision stump/tree results in the previous assignment?

As we used stochastic gradient descent, our convergence check checked if the difference between the weights from the start of one epoch (entire example set) to the end was less than $\epsilon = 0.01$ (determined from testing). In stochastic gradient descent, since you check one example at a time, it is important to not have the convergence check analyze any given $\frac{dL}{dW}$ because though the weight change may be small for that one example, it may not be so for the remaining examples. We summarize our results below

Data Set	AROC	Accuracy	Number of Iterations	Decision Stump Accuracy
Voting	0.9956	0.9886	15,800	0.9840
Volcanoes	0.8872	0.7490	52,800	0.7332
Spam	0.7070	0.6870	896,835	0.6639

These results show that the 0 hidden layer ANN outperforms the decision stump on all data sets. This is not very surprising because the ANN was allowed to grow to convergence. It is important to remember that the perceptron is not a very complex classifier though and weaker results are to be expected. This is because the perceptron compared to the ANN has many less parameters and fewer opportunities for nonlinearity. The ANN can apply nonlinearity to its computation at the output of every neuron, so only having 1 output neuron versus a layer of hidden neurons and the output produces far simpler results. This is reflected in later analysis where we explore the benefit of adding hidden neurons.

For volcanoes and spam explore how the AROC changes as learning iterations are increased. Fix the number of hidden units to 5 and =0.01 for these experiments. Plot AROC results for at least three values of learning iterations evenly spaced between 100 and 10,000. Compare your results to the "perceptron" in part (a). Does the introduction of hidden units lead to improved accuracy? How many iterations does this ANN need to converge compared to the "perceptron"?

Volcanoes

Number of Iterations	AROC	Accuracy
3000	0.7509	0.6726
6000	0.8269	0.6726
9000	0.8449	0.6726

Spam

Number of Iterations	AROC	Accuracy
3000	0.7377	0.6254
6000	0.7378	0.6254
9000	0.7428	0.6254

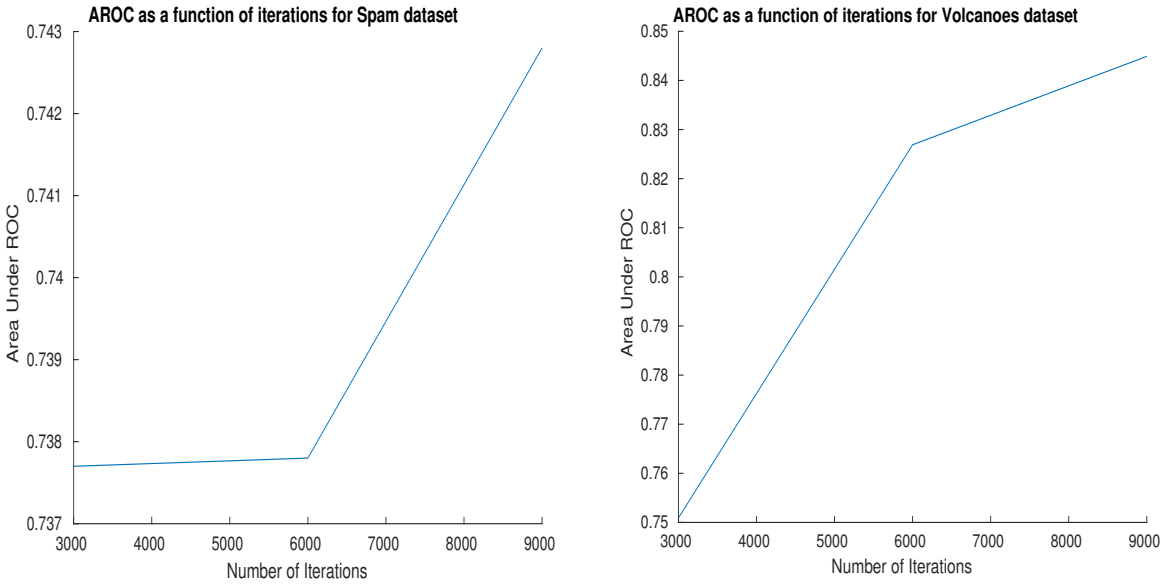

Figure 1: Plots of how AROC varies as a function of number of iterations for spam (Left) and volcanoes (right)

For the volcanoes data set, we see that the ROC graphs are nearly identical after only 9,000 iterations instead of 50,000 it needed to converge without hidden units. And for the spam data set, after only 9,000 iterations it actually outperformed the stump's AROC, though it still had a lower accuracy. We thus confirm that adding hidden units drastically reduces time to convergence for the reasons described in part a.

Second it is interesting to observe that although the accuracy is constant, the area under the ROC still changes. This is because accuracy is computed assuming that greater than 50% confident implies positive and less than 50% implies negative. Therefore, as confidences change between 50%-100% or 0%-49%, the accuracy measures won't change but the ROC graph will. The area under the ROC graph is maximized when the classifer is more confident, i.e the negatives are closer to 0 and the positives closer to 1 because then it has steeper edges. As the ANN trains it becomes more accurate and increases the ROC but does not effect the accuracy.

Explore how the AROC changes as the number of hidden units are varied on volcanoes. Plot AROC results for at least three values of hidden units evenly spaced between 3 and f, where f is the number of input units. Set =0.01 and the learning iterations to 100f , with a minimum of 10,000. Compare the ROC and training times to the results in parts (a) and (b). Warning: This experiment may take a long time to run.

Number of Hidden Units	AROC	Accuracy
75	0.8099	0.7399
150	0.8716	0.7803
225	0.8703	0.8386

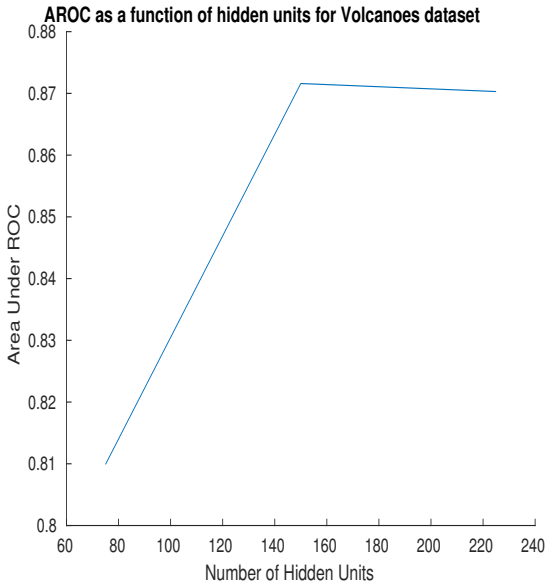

Figure 2: Plots of how AROC varies as a function of number of hidden units for volcanoes

Observe we see similar results as from experiment 2. We see that increasing the number of hidden units improves the performance. We see better performance in all trials of this test than for all trials with 5 hidden units. It does seem that the ANN reaches its peak performance after 150 hidden units as increasing both number of iterations and number of hidden units does not improve performance. This suggests that much like the perceptron is limited in how complex decision boundaries it could capture, so too is a 1 hidden layer

ANN and perhaps adding another layer is the best way to improve performance (though risking overfitting).