

Assignment No. 2: Mushroom Classification using BP-ANN

Given a dataset of 8124 hypothetical samples of mushrooms which included 23 different characteristics (for each sample), the goal of this assignment was to apply the technique of Back Propagation on a neural network to correctly classify a mushroom as poisonous or edible based on certain characteristics. The idea of BP is to tune the weights of the network so that our input gives us the desired output. This is done by changing the weights in such a way that we are minimizing the error function of our network.

One of the approaches i tried was to train the NN was to go through the entire training set applying BP each time it was needed and doing that iteratively until the *accuracy* of the epoch did not change. I also tried out continuously running through the set until the accuracy reached a certain threshold, but the problem was that sometimes that accuracy would plateau and not be able to increase after more back propagations. experimented with changing the the number of hidden neurons, number of output neurons, and the learning speed but was not able to find that a certain set of parameters was much better than another. I did however find that when i set the max epochs variable to a really large number like 1000 it had a negative impact on the result of testing the neural net. Below are some results.

Variables: maxEpochs = 100, number of hidden neurons = 16, number of output neurons = 1, learning speed = .1

Avg CPU run time : 2 seconds

Trial #	Accuracy (# mushrooms correctly classified/ # mushrooms)
1	.76
2	.86
3	.82
4	.27
5	.12
6	.79
7	.16
8	.28
9	.28
10	.18
Average accuracy	.45

If i were to try to make any other changes that could potentially improve my results i would try adding in another hidden layer and see how that would affect the result. Additionally i would do a more extensive analysis on finding parameters that would optimize my results.

references : Data Source 1: UCI ML Repository <https://www.kaggle.com/uciml/datasets>

\*accuracy is defined as the fraction of times the NN made the correct classification.



