Report: A1 Liner Classification

Name: Brandon Maness

Email: [bmaness3@uncc.edu](mailto:bmaness3@uncc.edu)

UNCC ID: 800828979

Problem Definition: This assignment tasks me with classifying a mushroom into one of two classes, poisonous or non-poisonous , based on the features of that mushroom.

Method: I did a brief review of the data then went to work converting the features to numbers instead of letters and normalizing the values that were converted. Then I used a perceptron model to train the dataset, updating the weights based on w = w + learningRate\*error\*current.

This yielded decently accurate results, I altered the iterations and learning rate multiple times before settling on 65 iterations to fit the data because at that point, there was no significant change to the error.

Dataset: The dataset is a 8124 rows by 23 columns set. The training set was the entirety of the dataset and the testing set was the class column of the dataset. I can make a list of features but that seems exhaustive as they are very self-explanatory, so I’ll define some of the terms used to clarify:

Class: This is either p for poisonous or n for non poisonous

Cap variables: These describe the rounded top of the mushroom and its characteristics

Bruises: These are small discolored spots that develop on the mushrooms naturally over time

Odor: This describes the smell of the mushroom

Gill variables: These attributes describe the undersides of the cap, which surround the stem

Stalk variables: These describe the ground-rooted stalk of the mushroom

Veil type/color: these describe the webbed ‘veil’ that some mushrooms produce at the edges of their caps that drape to the ground

Experimental Details: I’m using a linear classification algorithm taking in a single row of mushroom features and a bias to output a prediction for what class of mushroom the row belongs to. The data was given to me from the zip file provided for the project. I converted the features to numbers using a label encoder, this allowed me to treat the information as ranges of numbers instead of lets describing attributes. I didn’t add data to the dataset at all. I used an error calculation to evaluate how skewed my predictions were from the actual results, it seemed to work decently well.