Jacob Zahn and Dylan Scott

Professor Hamed Sari-Sarraf

ECE 4332

April 2, 2019

Project 4 Report

1. We used a permutation of our dataset to separate Testing and Training data. 20% of our dataset is used to test and 80% of our data is used to train. When tested our model had an accuracy of 93+%, meaning it only misidentified 7% of the samples it had never seen before.
2. Our dataset is in a .csv file named mushrooms. It contains the independent characteristics of 8125 mushroom species. These characteristics are as follows (class, cap-shape, cap-surface, cap-color, bruises, odor, gill-attachment, gill-spacing, gill-size, gill-color, stalk-shape, stalk-root, stalk-surface-above-ring, stalk-surface-below-ring, stalk-color-above-ring, stalk-color-below-ring, veil-type, veil-color, ring-number, ring-type, spore-print-color, population, habitat). For each sample and characteristic, a letter is used to denote an answer. These letters are translated to Integers so that the data is usable by the model. The class characteristic was either e or p denoting edible or poisonous respectively. We used the class characteristic as our targets, and all the other characteristics as the Xs. The goal for our model was to identify based on descriptions of characteristics whether a given mushroom was poisonous.