OMSCS\_CS7641 HW1

Clement Li (cli620) Fall 2020

# Introduction:

The purpose of this project is to explore some techniques in supervised learning. These techniques include: decision tree with pruning, neural networks, boosting on decision trees, support vector machines, and k-nearest neighbors. The first data set (CHECK) are a set of patients with 16 fields describing the patient and labeled on whether they have diabetes or not. The second data set are (CHECK) 128 by 128 by 3 images split up into their folders labeled with inhabitants of the coral reef seafloor.

# Dataset 1 – Diabetes:

## About/Preprocessing:

This dataset, pulled down from (CHECK) , each row represents an individual patient, whether they have any of the 16 characteristics of diabetes and a flag of if they were diagnosed with diabetes. Some example characteristics includes age, gender, obese, etc. There are 521 patients in this dataset. Besides the age field, the data is mostly binary; this would make the dataset as not as interesting. To make the problem more interesting, as part of the preprocessing phase,

## Decision Trees:

### Figures:

### Analysis:

## Neural Networks:

### Figures:

### Analysis:

## Boosting:

### Figures:

### Analysis:

## Support Vector Machines:

### Figures:

### Analysis:

## K-nearest neighbors:

### Figures:

### Analysis:

## Overall:

# Dataset 2 – Coral Reef:

## About/Preprocessing:

This dataset, pulled down from (CHECK) , is a set of images categorized into five different types of images seen on the coral reef ocean floors: brain corals, branching, favids, sand pavements and urchins. There are 72 brain corals, 49 branchings, 89 favids, 80 sand pavements and 14 urchins. Each 128 by 128 by 3 RGB image is read in, flattened to an 49152 by 1 array, and appended by the label value. This data is interesting because the specimens on the seafloor are normally very tightly clustered and images provided often contains other specimens. The specimens are unique in patterns, but majority of the images are often similar in color and texture. This provides an interesting problem in how would each supervised learning algorithm performs for image recognition of the coral seafloor.

https://www.narcis.nl/dataset/RecordID/oai%3Aeasy.dans.knaw.nl%3Aeasy-dataset%3A75772

## Decision Trees:

### Figures:

### Analysis:

## Neural Networks:

### Figures:

### Analysis:

## Boosting:

### Figures:

### Analysis:

## Support Vector Machines:

### Figures:

### Analysis:

## K-nearest neighbors:

### Figures:

### Analysis:

## Overall: