Project Proposal for CS 184A/284A, Fall 2019

Project Title: Classifying Heart Disease using UCI's Heart Disease Data Set

List of Team Members:
Kashan Saeed, 69974920, khsaeed@uci.edu

1. Project Summary

This project will use a neural network to predict the type of heart disease using the UCI Heart Disease Data Set, evaluating it using test data.

2. Problem Definition

The inputs will be features like age, sex, fasting blood sugar, maximum heart rate, and etc and the output will be which class the algorithm decides it falls into. Previously logistic regression and KNN have been used, both performing very well with accuracy's around 90 percent.

3. Proposed Technical Approach

The system I am using is not one with multiple pipelines or components, that because the dataset I am using is just a bunch of values. I plan to use neural networks to see if it would give more accurate results than linear regression. I plan to try out many different activation functions to see how they will effect the results.

4. Data Sets

The dataset I am using is located here https://archive.ics.uci.edu/ml/datasets/heart+Disease , it has five classes and 76 features, though from reading I found out that most ML researchers use the same 14 features or more specifically the "Cleveland Database".

5. Experiments and Evaluation

I plan to use k fold cross validation to try out different models, and to keep some test data on the side for the model that I decide to finalize with. It does not seem like the data comes with partitions so I will just divide the data before training the models.

6. Software

1. Python, Jupyter notebooks, numpy, pandas, tensorflow
2. I plan to view the data, do any cleaning that may be necessary, feature scaling, and training and testing

7. Milestones

- Weeks 7 and 8
    - Collect and analyze data
    - Figure out how to handle NaN's
    - Build Neural Nets

- Weeks 9 and 10
    - Continue building neural nets, and testing different configurations
    - Writing proposal

8. Individual Student Responsibilities

Kashan: Will write and test the code, collect data, write report. Basically do everything.