Heart Disease Prediction Using Adaptive Network-Based Fuzzy Inference System (ANFIS)

Erin M. Kunz, Department of Electrical Engineering

CS 229 Spring 2019, Stanford University

Introduction

Heart Disease is the leading cause of death for both men and women in the United States. Approximately 610,000 people die from heart disease each year, which is equivalent to 1 in every 4 deaths. This study attempts to improve prediction performance past that of existing prediction systems to further improve and save lives. The input will be 10-13 features of medical data that are easily obtainable and typically used for heart disease diagnosis and the output will be the prediction of heart disease. A variety of traditional classification algorithms will be ran and tuned. Furthermore, an Adaptive Network-based Fuzzy Inference System will be trained and tested.

Data and Features

- Two datasets from the UCI Machine Learning Repository [1] were analyzed, for a total of 760 subjects, of which 348 had the diagnosis of heart disease.
- Features analyzed included:
  - Age
  - Sex
  - Chest Pain Type
  - Resting Blood Pressure
  - Cholesterol
  - Fasting Blood Sugar
  - Resting electrocardiographic results
  - Exercise Induced Angina
  - ST Depression Induced by Exercise Relative to Rest
  - Slope of the Peak Exercise ST Segment
  - Number of Major Vessels Colored by Flourosopy
  - Presence of Defect, Reversible or Fixed
- Data was taken from studies in Switzerland, Hungary and Cleveland, Ohio.

Figure 1: Correlation matrix for 10 features for the combined datasets (left), correlation matrix for all 13 features for the first dataset (right).
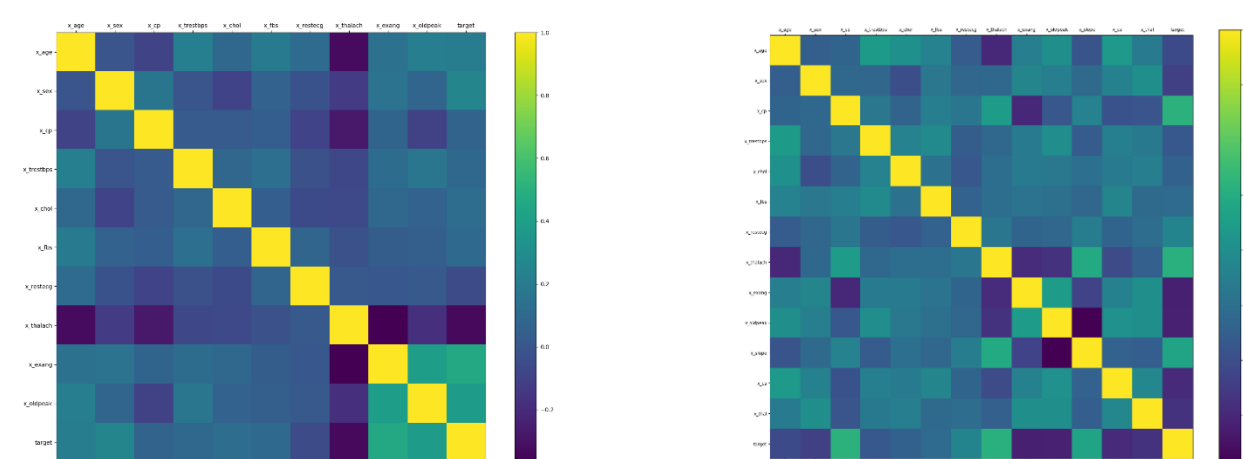
Models

- Logistic regression is a nonlinear transformation of the linear regression model between the input variables and the binary class assignment. It maps this linear regression using a function like the sigmoid (S-shape) so that all probabilities are mapped between 0 and 1.
- In KNN, the K nearest neighbors of a given data point are analyzed and the majority class of these K neighbors is assigned to the data point.
- Support vector machines is a machine learning technique that looks at separations in data when mapped to higher-dimensional space and determines which side of the split a data point is in.
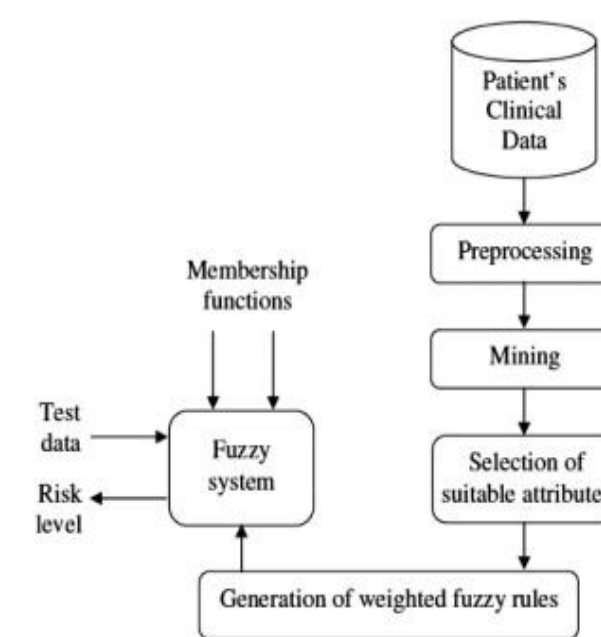
- ANFIS combines the structure of a neural network with "fuzzy logic" principles that attempt to replicate the nonlinearities of a medical professional's experiential knowledge when making a decision about a patients diagnosis based on medical data. This principle of fuzzy logic looks at a classification more as a sliding scale between 0 and 1, where there may be more of a continuity to a subject's diagnosis of heart disease instead of a cut and dry binary classification.

Results

- The highest accuracy of 89% was achieved using a KNN classifier with K = 10 and ANFIS with 24 membership rules and a learning rate of 0.07, both on the smaller dataset with three more features.
- The neural network classifier performed the worst.
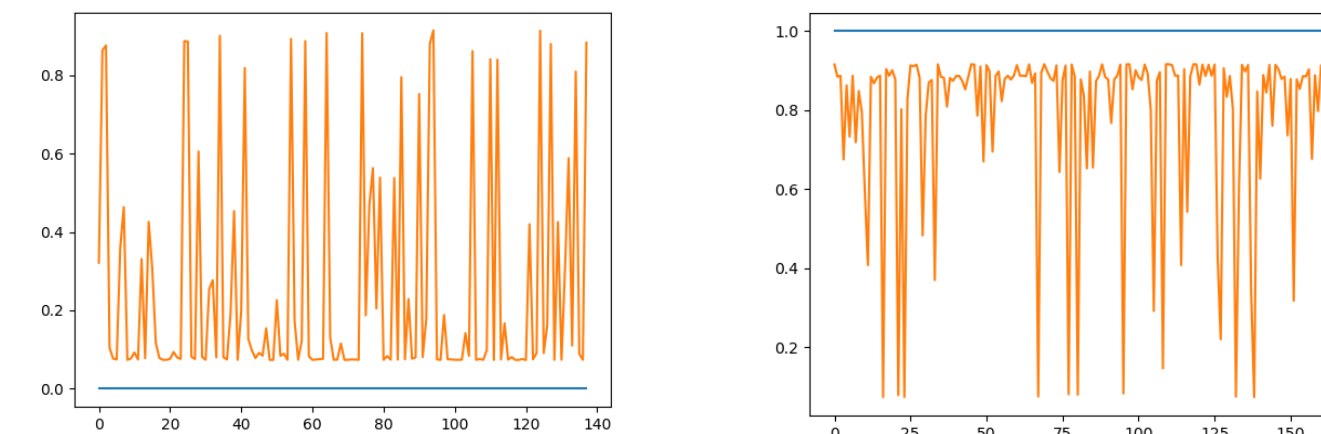
DATASET	LR	KNN	SVM	NN	ANFIS
SMALL W/ 13 FEATURES	0.809	0.89	0.84	0.787	0.891
LARGE W/ 10 FEATURES	0.836	0.87	0.86	0.803	0.820

Conclusions

- Well-tuned traditional classifier algorithms perform well in predicting heart disease
- ANFIS shows potential for improving upon traditional classifiers
- Feature space appears to be a limitation in further increasing accuracy

Future Work

- Obtain and analyze data with a greater feature space
- Look into combing fuzzy logic principles with a higher performing traditional classifier, such as KNN or SVM
- Utilize a medical professional's experiential knowledge to initialize membership functions in ANFIS

References

[1] Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

[2] Ibrahim, R., Olawale, O. and Funmilayo, K. (2018). Diagnosis of Hepatitis using Adaptive Neuro-Fuzzy Inference System (ANFIS). *International Journal of Computer Applications*, 180(38), pp.45-53.

[3] Dogantekin, E., Dogantekin, A., Avci, D. and Avci, L. (2010). An intelligent diagnosis system for diabetes on Linear Discriminant Analysis and Adaptive Network Based Fuzzy Inference System: LDA-ANFIS. *Digital Signal Processing*, 20(4), pp.1248-1255.

[4] Weng, S.F., Reps, J., Kai, J., Garibaldi, J.M., Qureshi, N. (2017) Can Machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS ONE*, 12 (4), art. no. 0174944.