NEURAL NETWORKS IN HEART DISEASE DETECTION

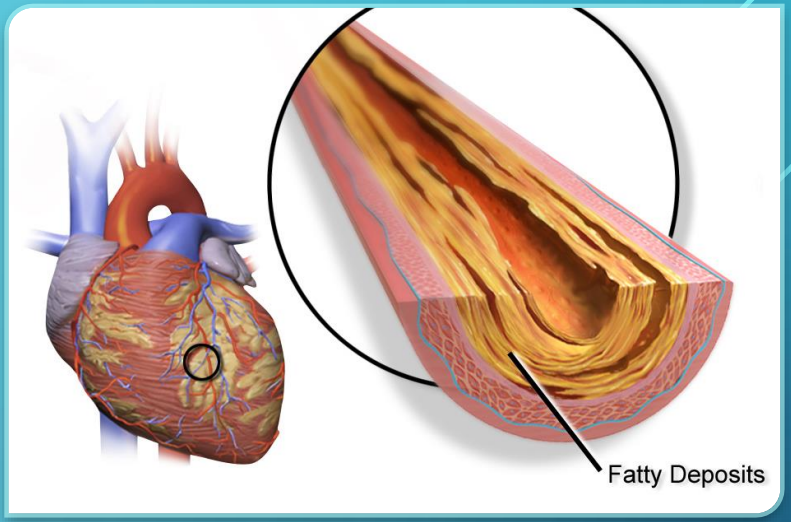
Rajesh thomas (rr15f)

Sourindu chatterjee (sc15)

MACHINE INTELLIGENCE

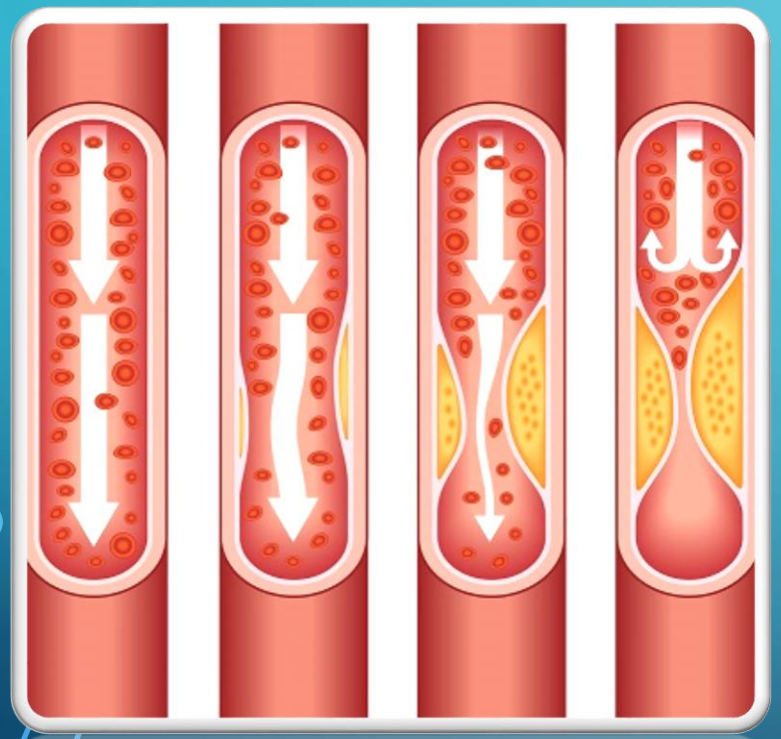
DR. SIMON FOO

INTRODUCTION



- Heart supplies blood to different parts of body
- Heart is a muscle (myocardium), so it itself needs blood to function
- Arteries that carry blood to the myocardium are called coronary arteries.
- Coronary arteries can get blocked, in general called Coronary Heart Disease (CHD)

MOTIVATION



- About 610,000 people die in the United States due to heart diseases every year.
- That's about 1 in every 4 deaths
- Coronary Heart Disease (CHD) is the most common type of heart disease.
- CHD causes more than 370,000 deaths annually.
- Modern Machine Learning algorithms, combined with powerful hardware have become sophisticated enough to aid CHD detection.

Heart Disease Death Rates, 2008-2010
Adults, Ages 35+, by County

Age-Adjusted
Average Annual
Rates per 100,000
- 122.5 - 306.2
- 306.3 - 350.9
- 351.0 - 391.6
- 391.7 - 451.9
- 452.0 - 846.1
- Insufficient Data

Alaska

Hawaii

Rates are spatially smoothed to enhance the stability of rates in counties with small populations.

Data Source:
National Vital Statistics System
National Center for Health Statistics

CDC

PROJECT GOALS

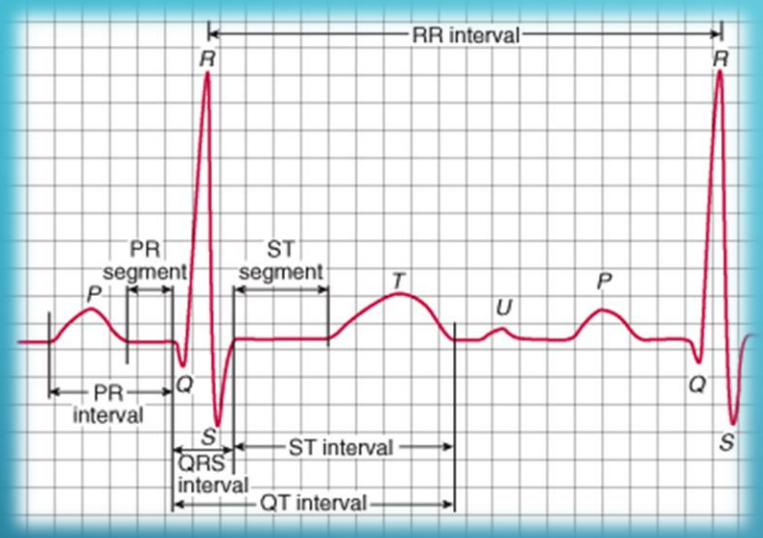
- Predict if a patient will get heart disease or not using the UCI database
- Preprocessing data using PCA
- Compare the performance of different algorithms
- Algorithms used are K-Nearest Neighbor, Backpropagation, Decision Tree and Radial Basis Function.

UCI DATABASE

- Very comprehensive database which contains data from different parts of the world.
- Provides many attributes which records many parameters that checks the health of the heart
- Provides ample number of samples

CHD

- Age
- Gender
- Chest Pain Type (4 types considered)
- Resting Blood Pressure
- Cholesterol
- Blood Sugar
- Resting ECG
- Maximum Heart Rate
- Exercise Induced Angina (chest pain)
- ST Depression Induced by Exercise

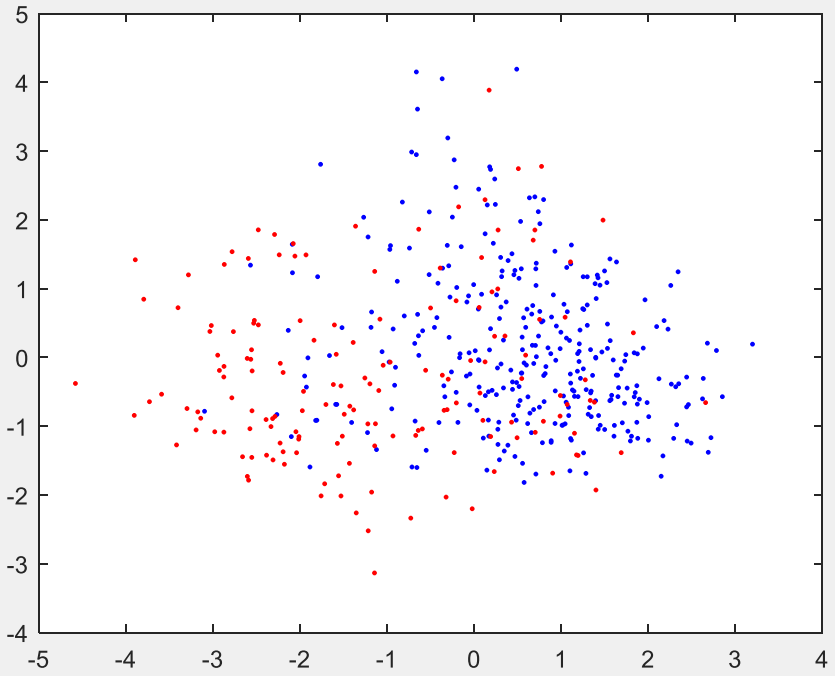


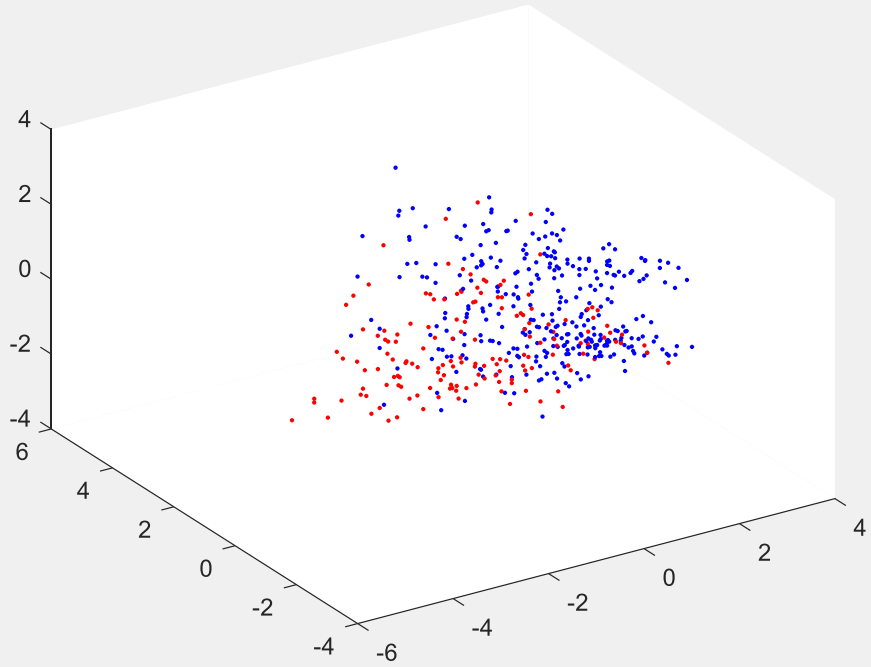
sphygmomanometer

PRINCIPAL COMPONENT ANALYSIS

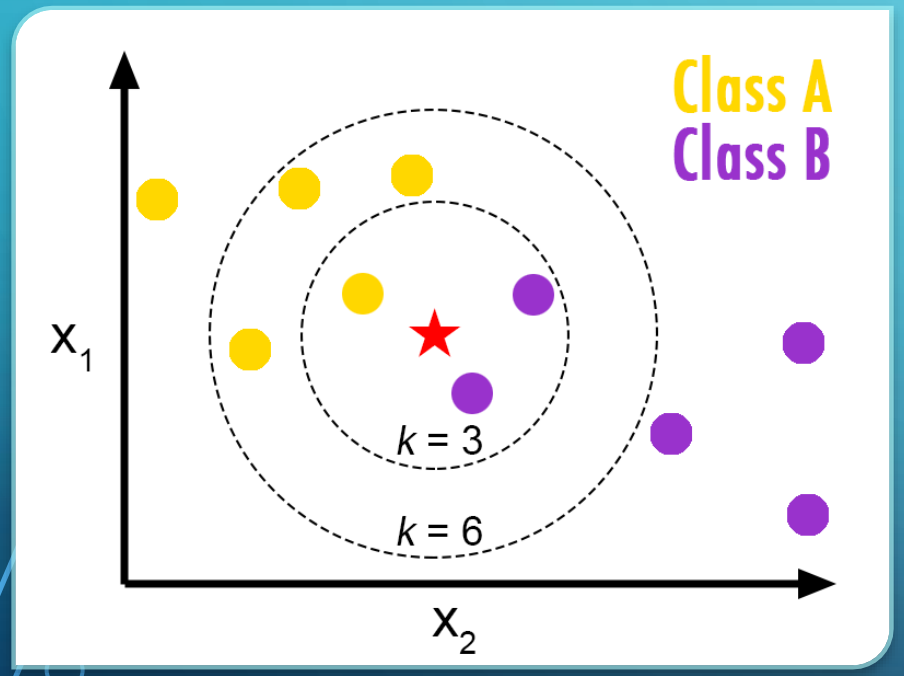
- Orthogonal transformation to convert correlated values into uncorrelated variables.
- Uncorrelated variables are called principal components
- This helps in feature reduction for better feature separation

PCA GRAPHS



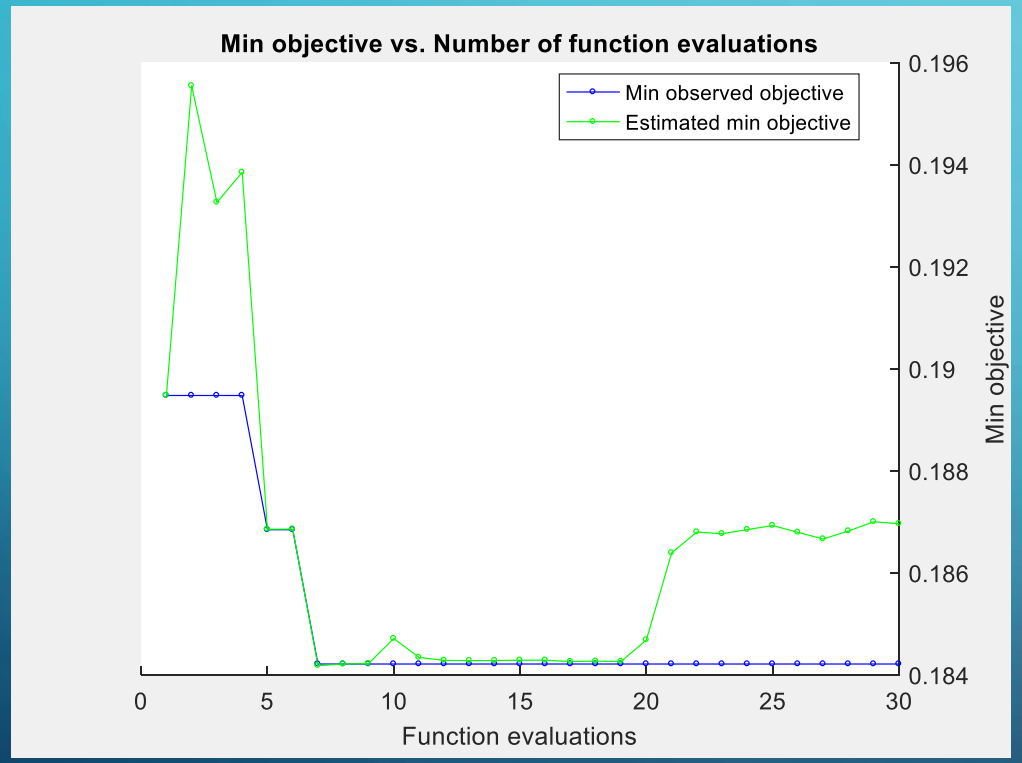


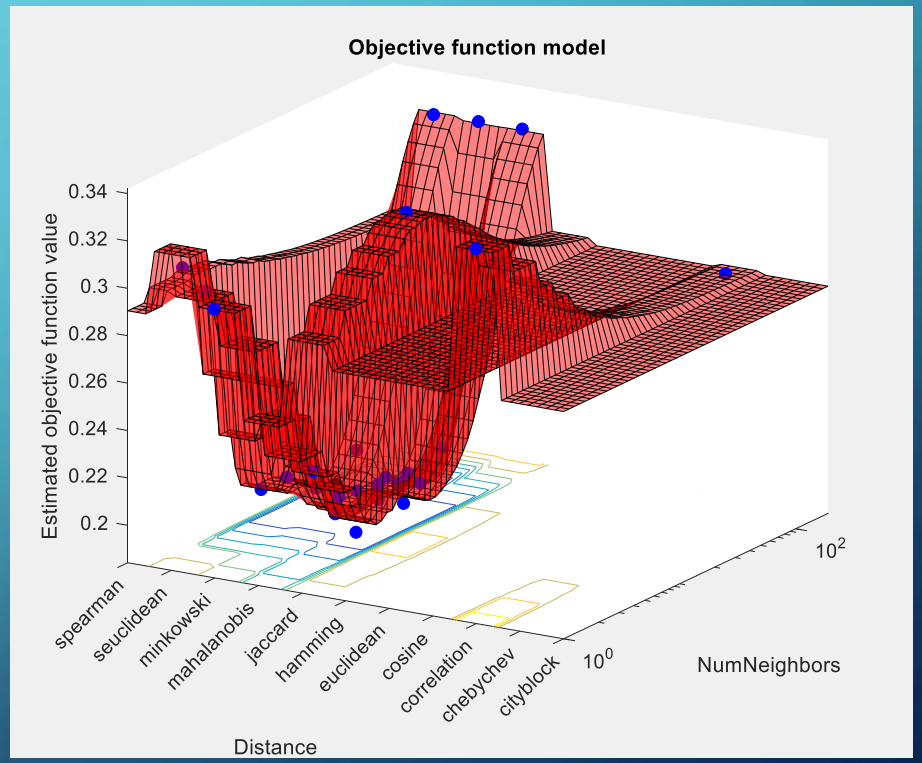
K-NEAREST NEIGHBOR



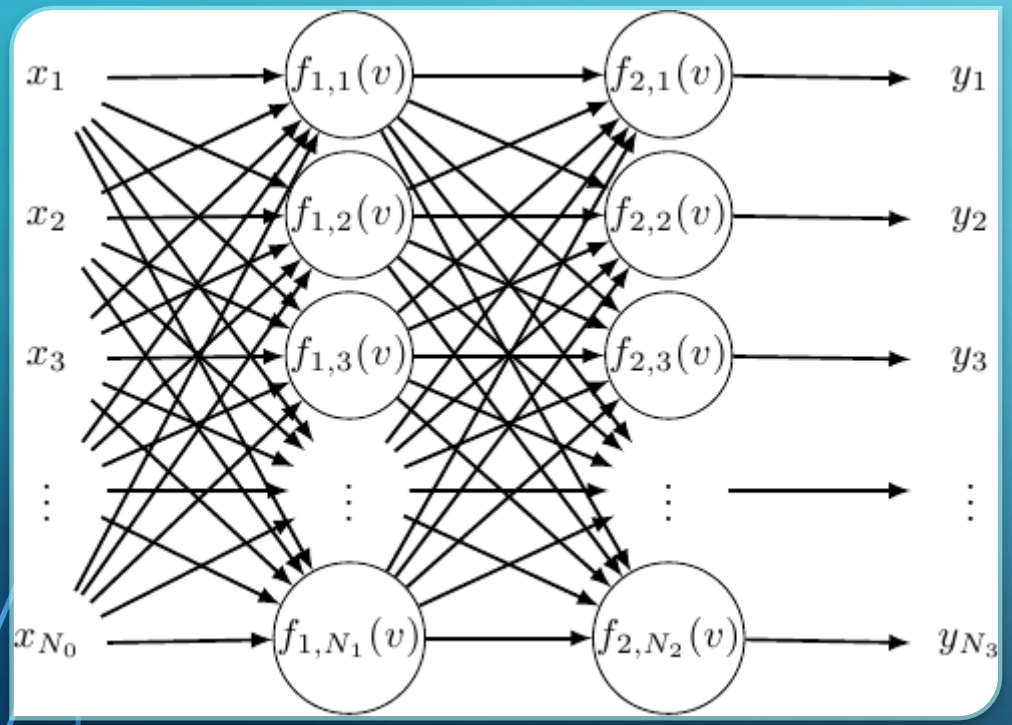
- Find the number of k nearest values corresponding to the new observation.
- Find which attribute has the maximum frequency among them.
- The predicted output will be the one with the maximum frequency
- The algorithm optimizes the value of k and the distance function to give the best results.

KNN OPTIMIZATION GRAPHS



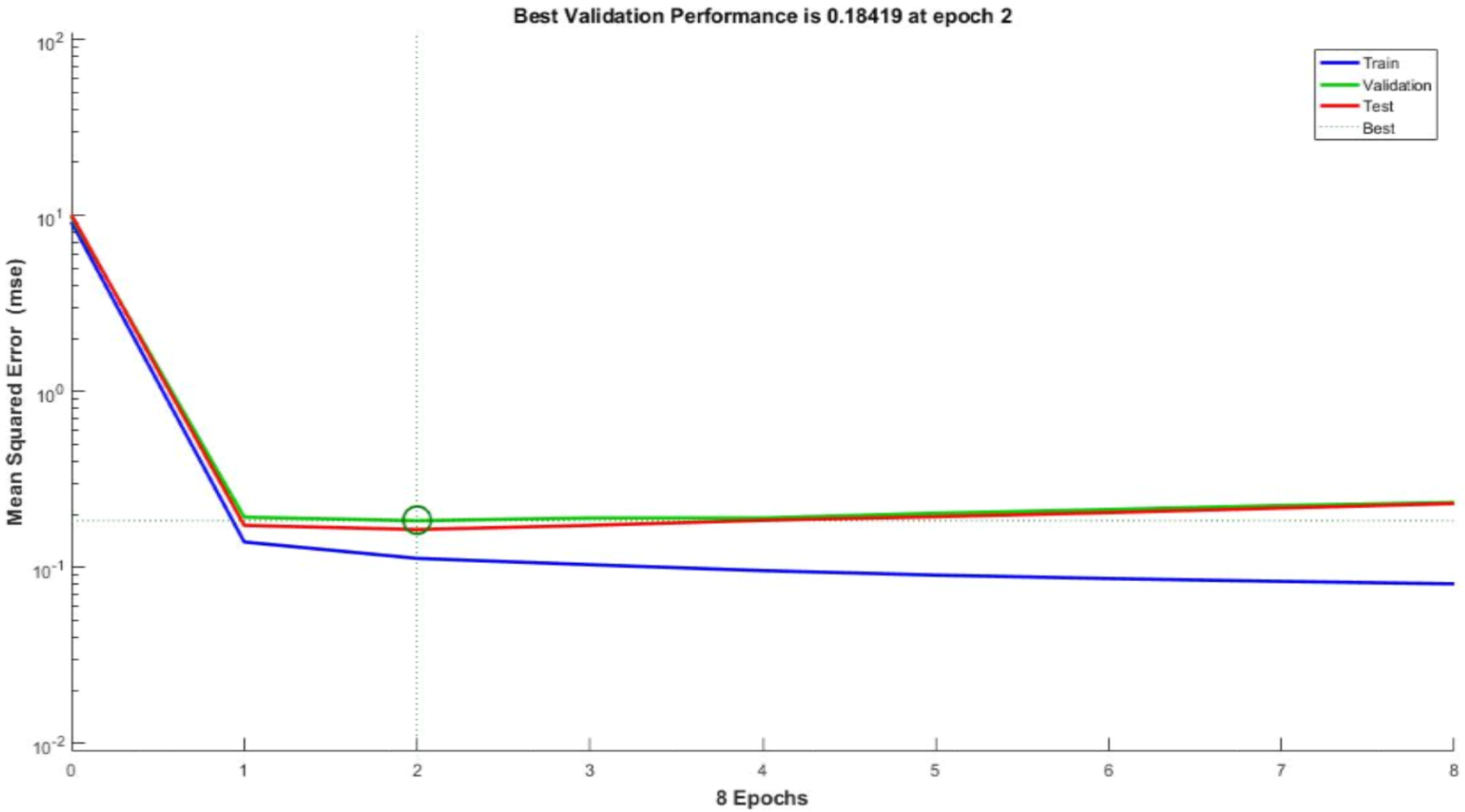


BACKPROPAGATION ALGORITHM

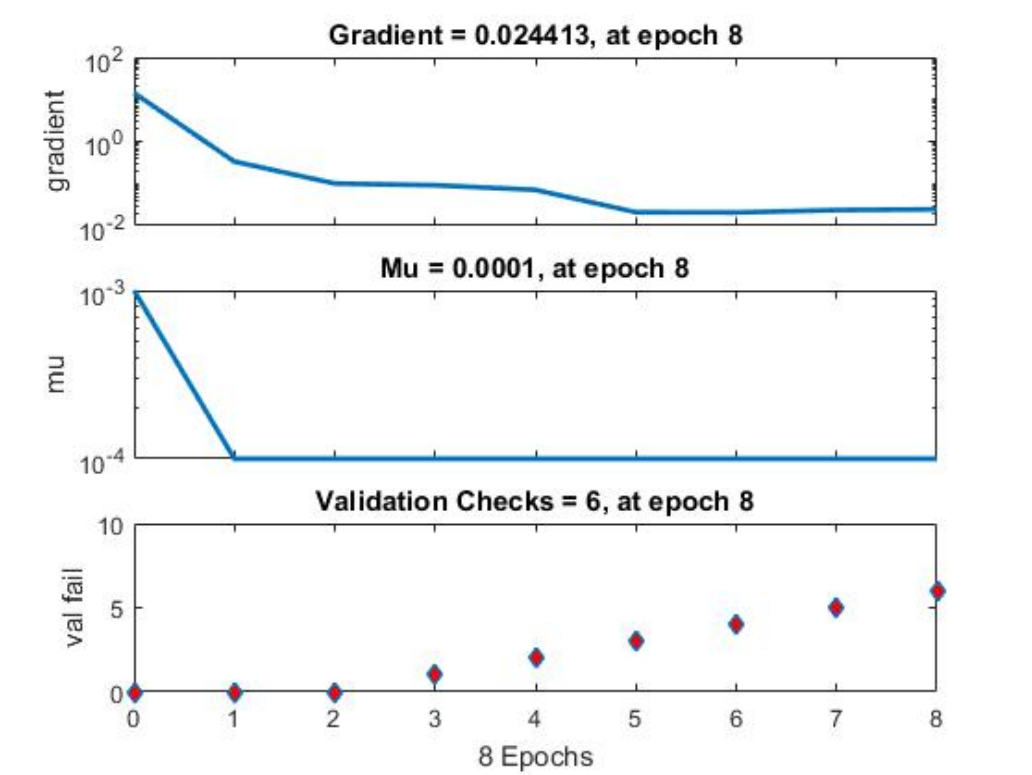


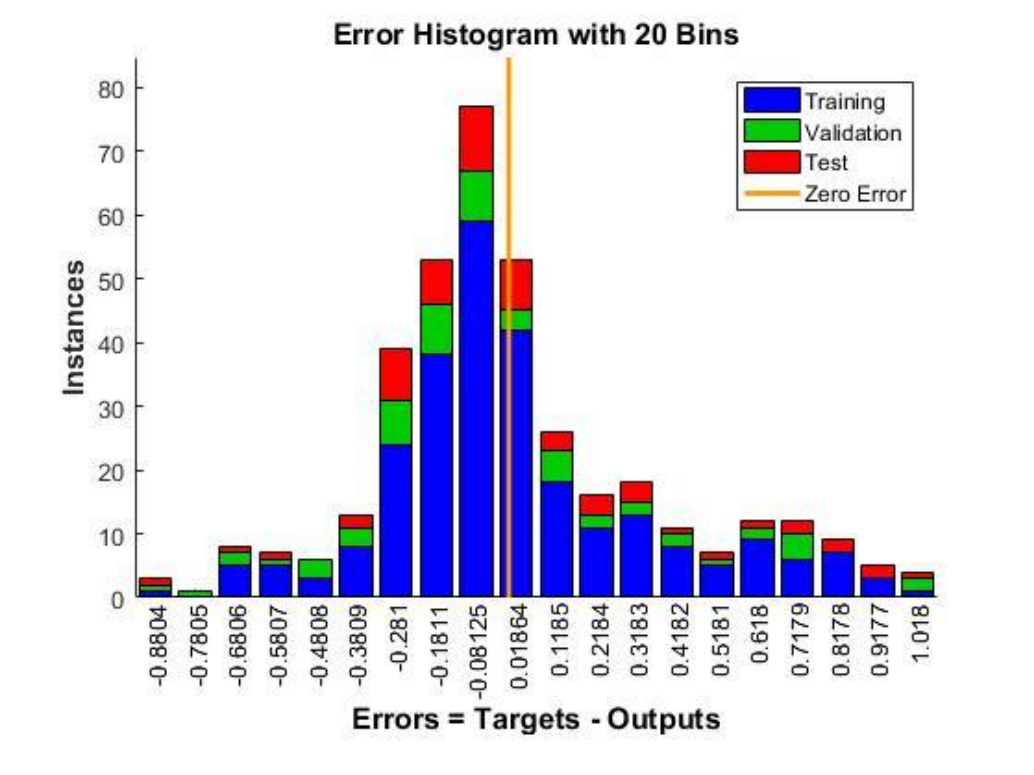
- The Multi-Layer Feed-Forward Network is effectively trained using Backpropagation Algorithm.
- Calculates the gradient of a error function e.g MSE, with respect to the weights in the network.
- The gradient is used to update the weights so that the error function is minimized.
- Outputs are calculated in both the layers using Sigmoid function.

BACKPROPAGATION TRAINING

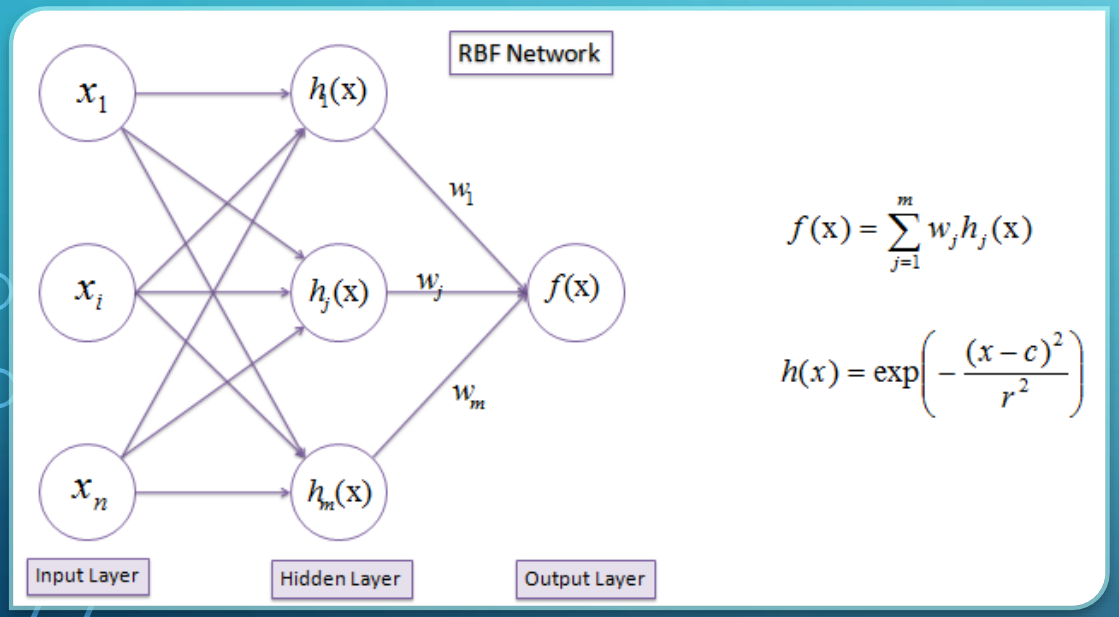


BACKPROPAGATION TRAINING





RADIAL BASIS NETWORK

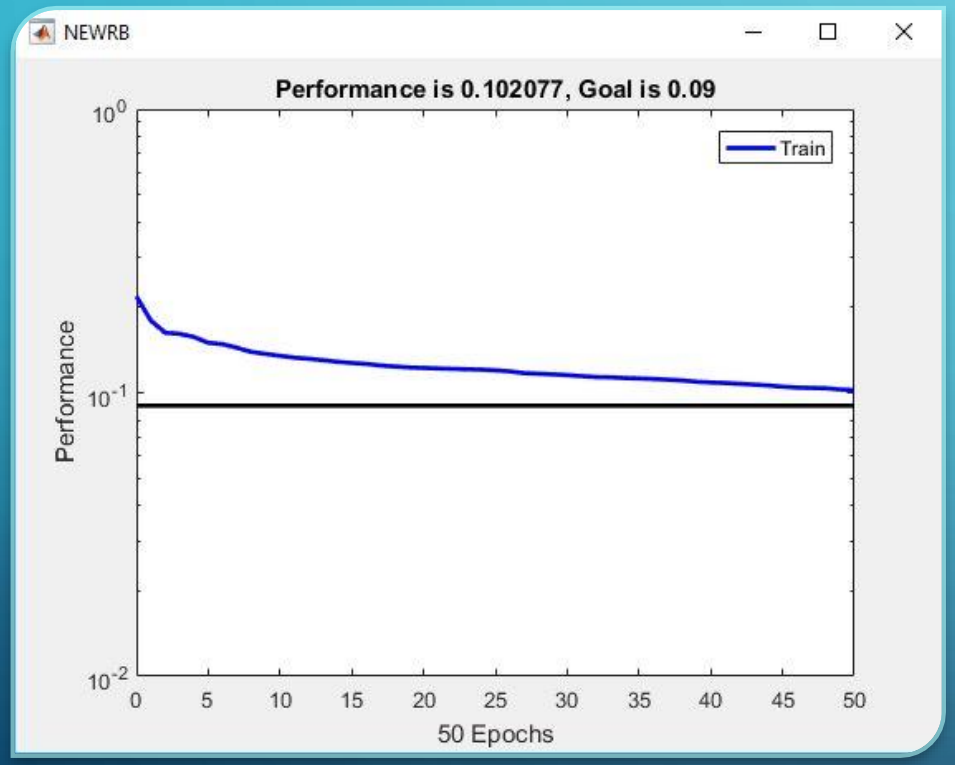


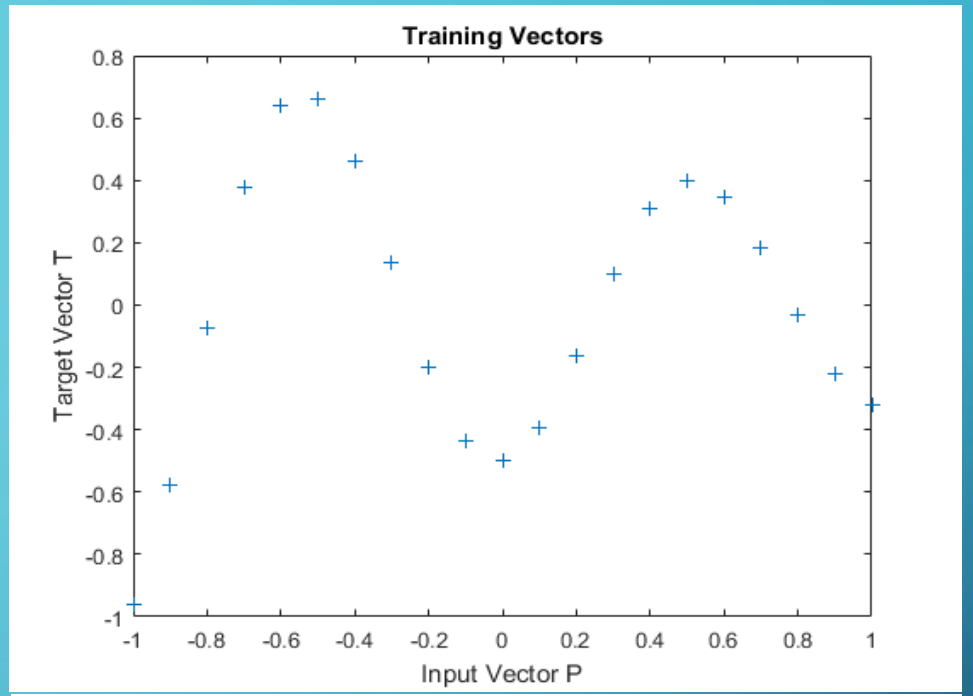
$$f(x) = \sum_{j=1}^{m} w_j h_j(x)$$

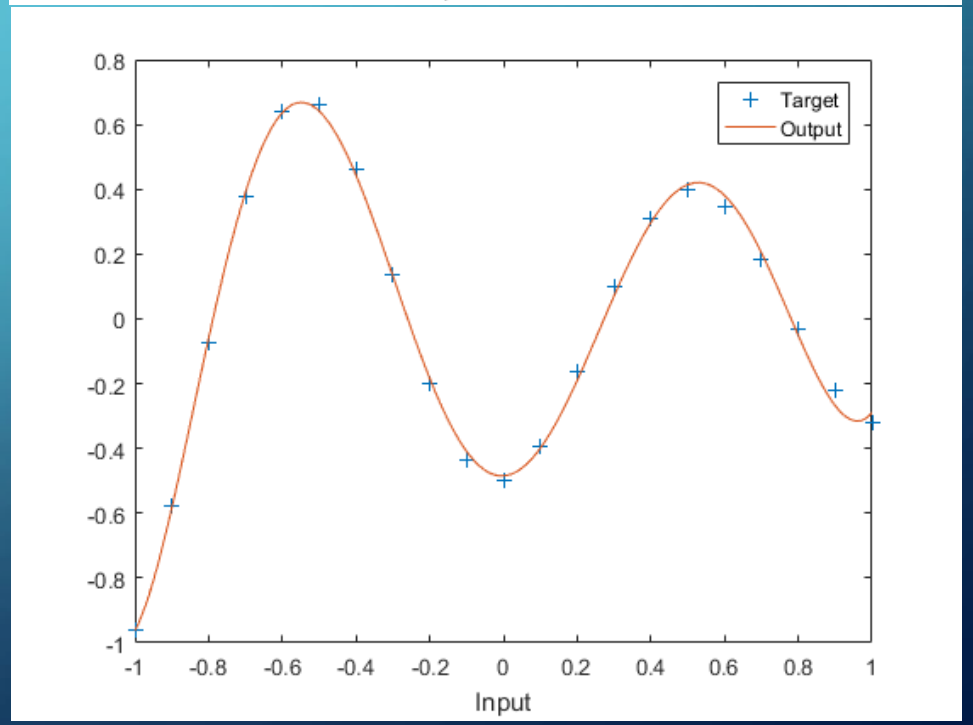
$$h(x) = \exp\left(-\frac{(x-c)^2}{r^2}\right)$$

- Works by approximating functions
- Approximation is done by iterative training on the given data
- The error on the training data can be reduced to values as small as desired
- New neurons can be added until the mean square error goal is achieved.

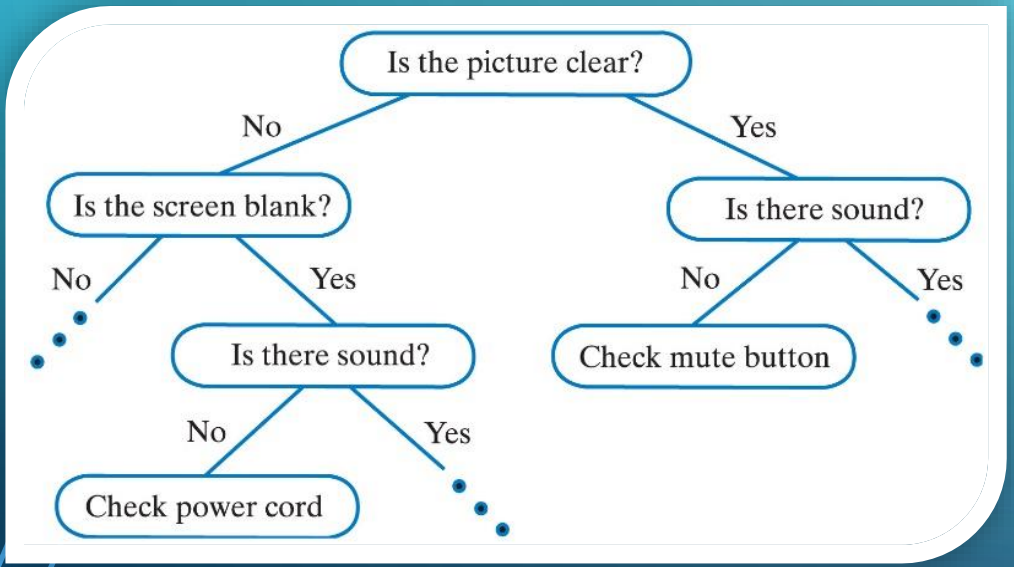
RBF TRAINING





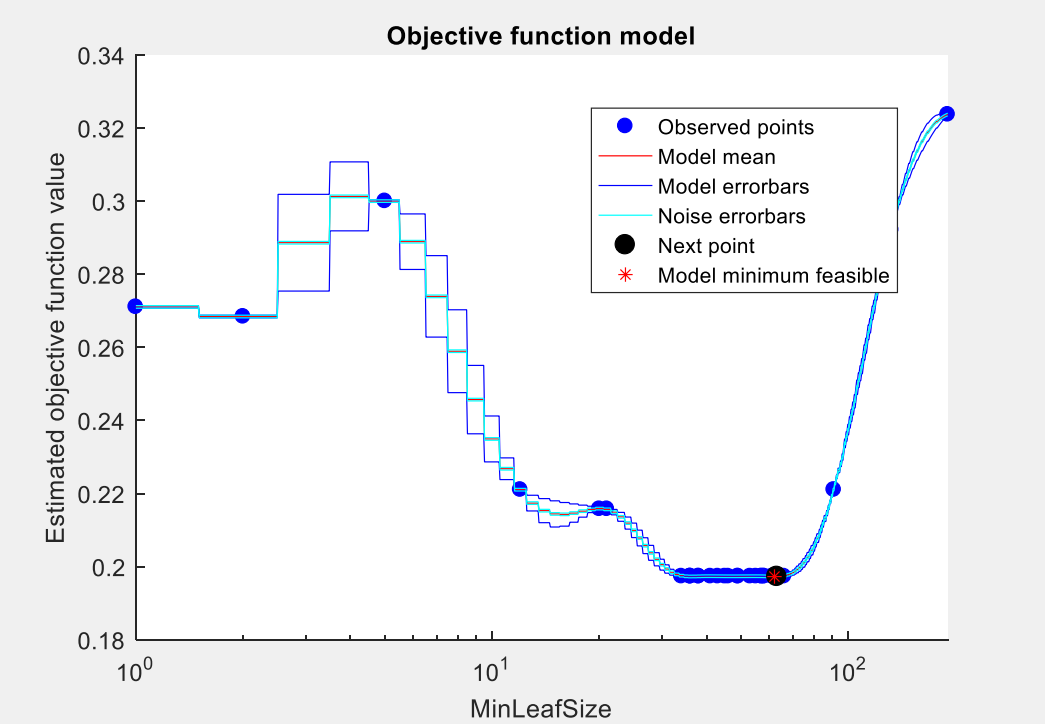


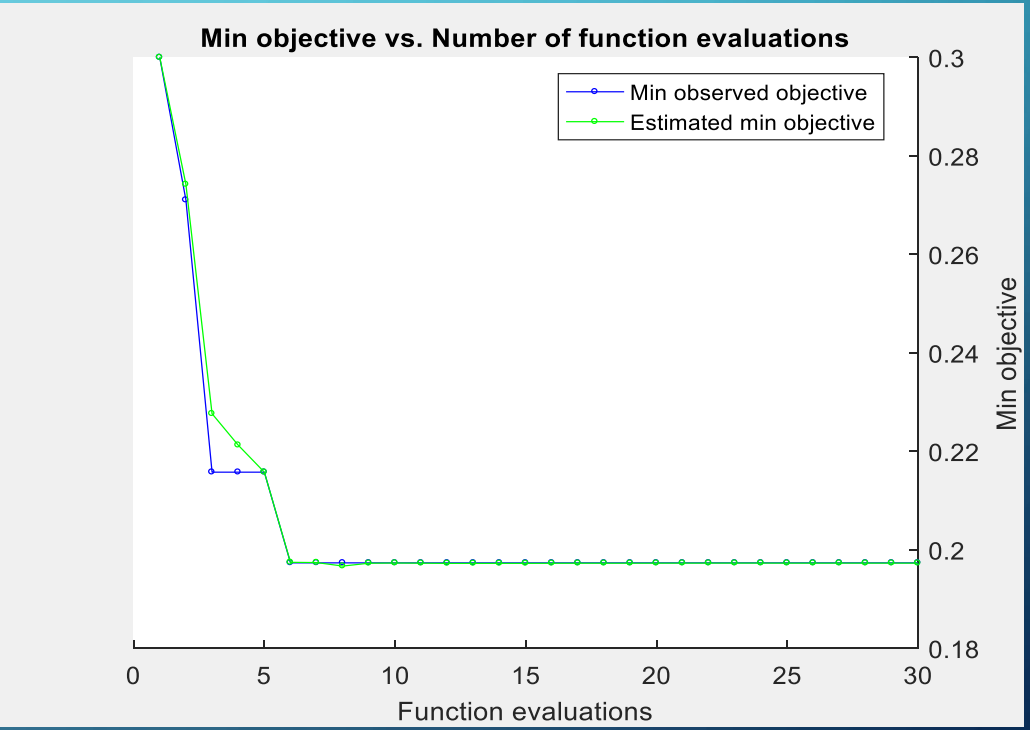
DECISION TREE



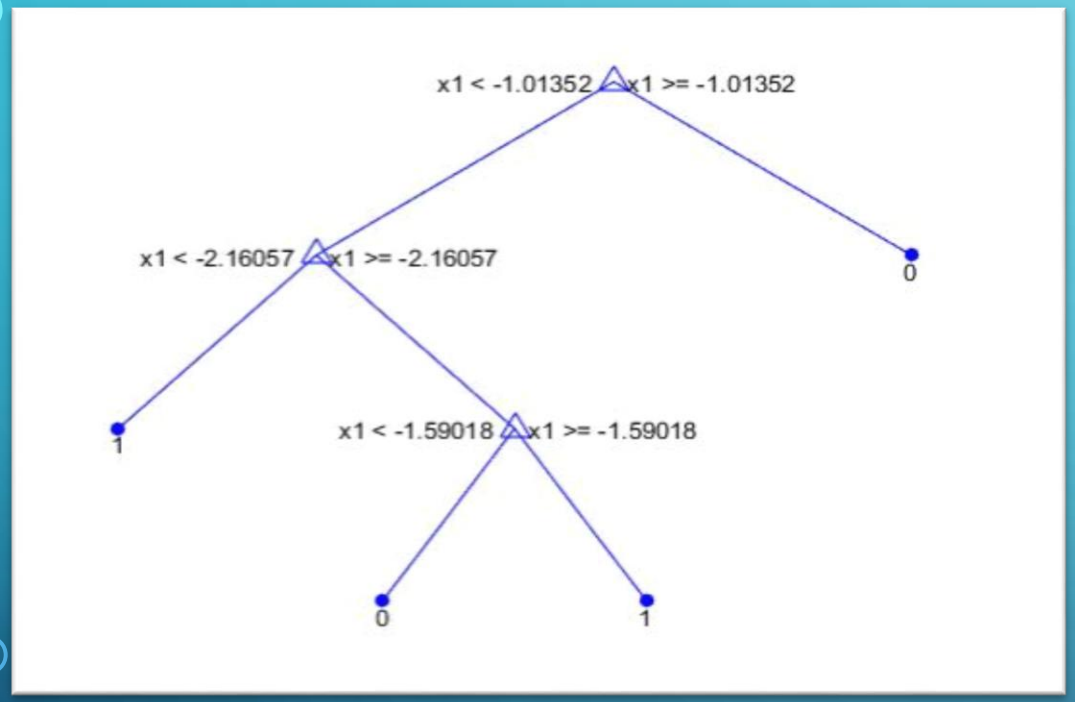
- Decision support tool with a flowchart like structure
- Traditionally created using hand
- Grow a decision tree based on the input values and a desired cost function like Entropy.
- Cross check with testing data

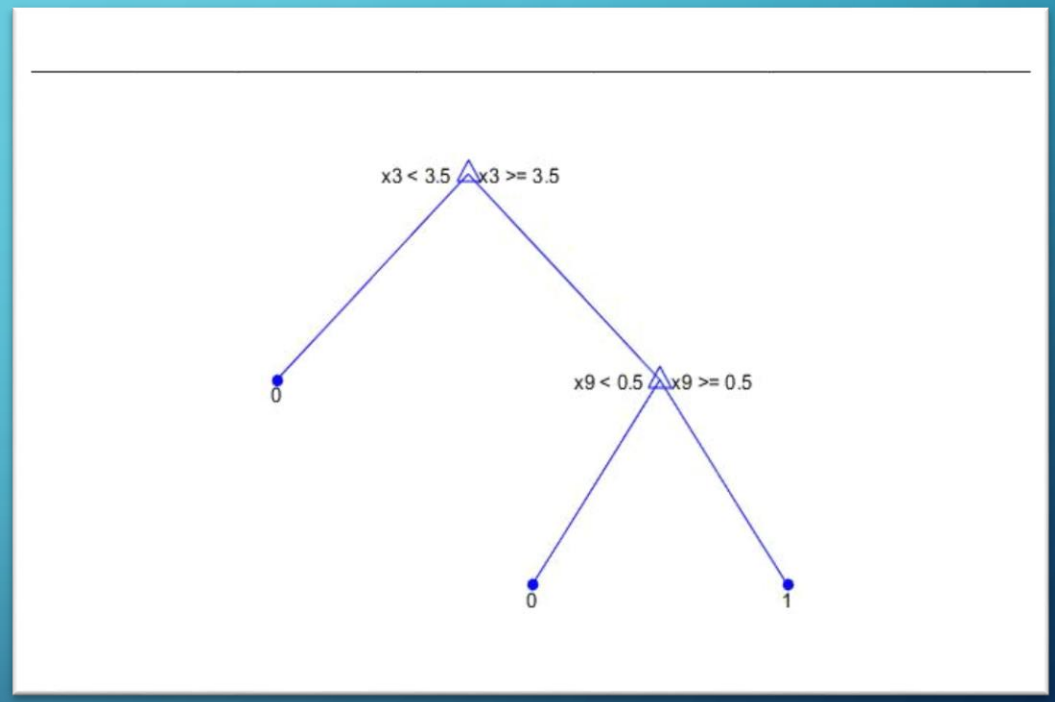
TREE OPTIMIZATION



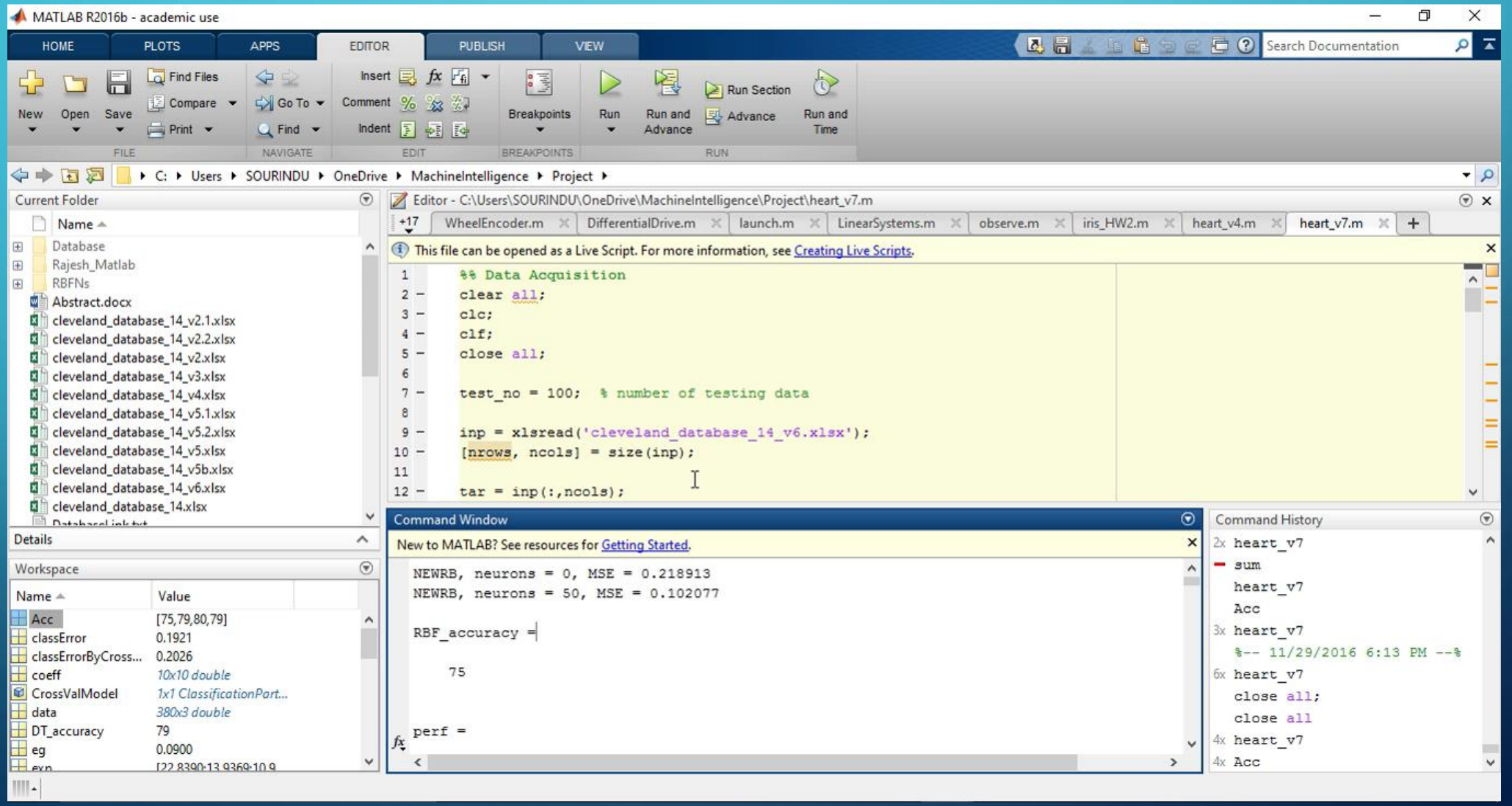


THE DECISION TREE DEVELOPED

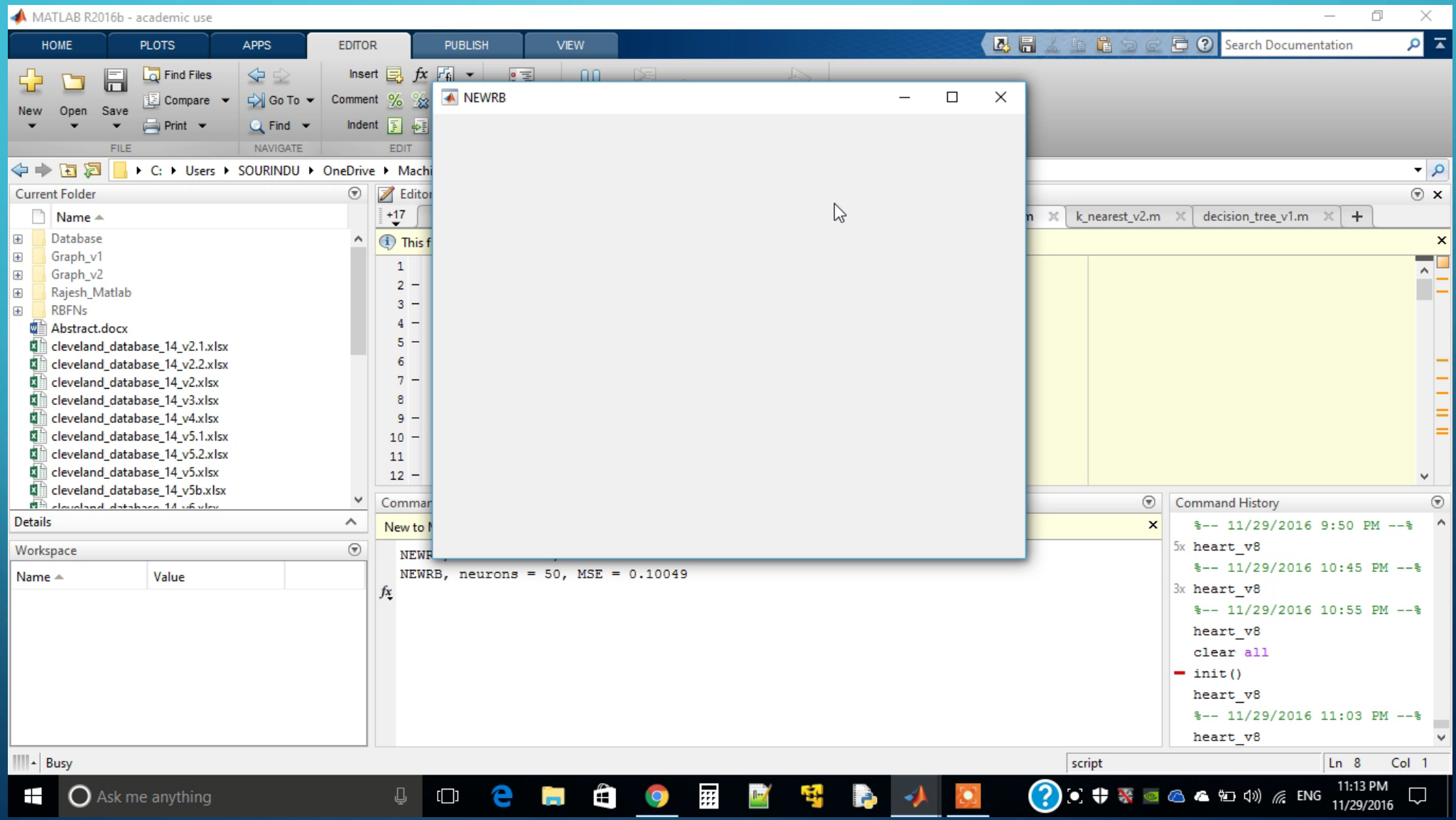




CODE VIDEO


MATLAB R2016b - academic use
HOME
PLOTS
APPS
EDITOR
PUBLISH
VIEW
Search Documentation
New
Open
Save
Find Files
Compare
Print
Go To
Find
Insert
Comment
Indent
Breakpoints
Run
Run and Advance
Run Section
Advance
Run and Time
FILE
NAVIGATE
EDIT
BREAKPOINTS
RUN
C: ▸ Users ▸ SOURINDU ▸ OneDrive ▸ MachineIntelligence ▸ Project ▸
Current Folder
Name
Database
Rajesh_Matlab
RBFNs
Abstract.docx
cleveland_database_14_v2.1.xlsx
cleveland_database_14_v2.2.xlsx
cleveland_database_14_v2.xlsx
cleveland_database_14_v3.xlsx
cleveland_database_14_v4.xlsx
cleveland_database_14_v5.1.xlsx
cleveland_database_14_v5.2.xlsx
cleveland_database_14_v5.xlsx
cleveland_database_14_v5b.xlsx
cleveland_database_14_v6.xlsx
cleveland_database_14.xlsx
Details
Workspace
Name
Value
Acc
[75,79,80,79]
classError
0.1921
classErrorByCross...
0.2026
coeff
10x10 double
CrossValModel
1x1 ClassificationPart...
data
380x3 double
DT_accuracy
79
eg
0.0900
Editor - C:\Users\SOURINDU\OneDrive\MachineIntelligence\Project\heart_v7.m
+17
WheelEncoder.m
DifferentialDrive.m
launch.m
LinearSystems.m
observe.m
iris_HW2.m
heart_v4.m
heart_v7.m
This file can be opened as a Live Script. For more information, see Creating Live Scripts.
%% Data Acquisition
clear all;
clc;
clf;
close all;
test_no = 100; % number of testing data
inp = xlsread('cleveland_database_14_v6.xlsx');
[nrows, ncols] = size(inp);
tar = inp(:,ncols);
Command Window
New to MATLAB? See resources for Getting Started.
NEWRB, neurons = 0, MSE = 0.218913
NEWRB, neurons = 50, MSE = 0.102077
RBF_accuracy =
75
perf =
Command History
heart_v7
sum
heart_v7
Acc
heart_v7
%-- 11/29/2016 6:13 PM --%
heart_v7
close all;
close all
heart_v7
Acc

TRAINING VIDEO



CONCLUSION

- Accuracy of about 85, 90, 89, 80 percent for RBF, MLFF (Backpropagation), KNN, Decision Tree.
- The development for an effective weight function to be used in PCA is in progress.
- The higher accuracy can be achieved using revised PCA to gain the maximum information.

FUTURE WORK

- Use more number of attributes given in the UCI database
- Apply the techniques used for a larger database
- Create a reliable phone app for real-time use.

QUESTIONS?