**Lab -30 July 2019**

**Two datasets are given Description about dataset is given below.**

1. **Examine the datasets**
2. **Do preprocessing ,check for missing values, normalize if required**
3. **Generate correlation matrix, heatmap**
4. **Do feature extraction**
5. **Implement KNN and SVM classification algorithms**
6. **Compare the performance**

Context-

dataset heartattack.csv

This database contains 76 attributes, but all published experiments refer to using a subset of 14 of them. In particular, the Cleveland database is the only one that has been used by ML researchers to this date. The "goal" field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4.

Content

Attribute Information:   
> 1. age   
> 2. sex   
> 3. chest pain type (4 values)   
> 4. resting blood pressure   
> 5. serum cholestoral in mg/dl   
> 6. fasting blood sugar > 120 mg/dl  
> 7. resting electrocardiographic results (values 0,1,2)  
> 8. maximum heart rate achieved   
> 9. exercise induced angina   
> 10. oldpeak = ST depression induced by exercise relative to rest   
> 11. the slope of the peak exercise ST segment   
> 12. number of major vessels (0-3) colored by flourosopy   
> 13. thal: 3 = normal; 6 = fixed defect; 7 = reversable defect

The names and social security numbers of the patients were recently removed from the database, replaced with dummy values. One file has been "processed", that one containing the Cleveland database. All four unprocessed files also exist in this directory.

To see Test Costs (donated by Peter Turney), please see the folder "Costs"

Inspiration

Experiments with the Cleveland database have concentrated on simply attempting to distinguish presence (values 1,2,3,4) from absence (value 0).

See if you can find any other trends in heart data to predict certain cardiovascular events or find any clear indications of heart health.

age

age in years

sex

(1 = male; 0 = female)

cp

chest pain type

trestbps

resting blood pressure (in mm Hg on admission to the hospital)

chol

serum cholestoral in mg/dl

fbs

(fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)

restecg

resting electrocardiographic results

thalach

maximum heart rate achieved

exang

exercise induced angina (1 = yes; 0 = no)

oldpeak

ST depression induced by exercise relative to rest

slope

the slope of the peak exercise ST segment

ca

number of major vessels (0-3) colored by flourosopy

thal

3 = normal; 6 = fixed defect; 7 = reversable defect

target

1 or 0

dataset weather.csv

This dataset contains weather data for New Delhi, India.

Content

This data was taken out from wunderground with the help of their easy to use api. It contains various features such as temperature, pressure, humidity, rain, precipitation,etc.

Inspiration

The main target is to develop a prediction model accurate enough for predicting the weather

Columns

* datetime\_utc
* \_conds
* \_dewptm
* dewpoint in C
* \_fog
* \_hail
* \_heatindexm
* \_hum
* humidity
* \_precipm
* \_pressurem
* \_rain
* \_snow
* \_tempm
* \_thunder
* \_tornado
* \_vism
* \_wdird
* \_wdire
* \_wgustm