

# Data Science Techniques and Applications

## Coursework I

Baran Buluttekın

March 8, 2019

### Phase I

For this coursework after some searching I choose to examine Heart Disease UCI dataset in kaggle. It comprises of data that collected from patients from 4 medical institutions. There is no timestamp associated with data collections but according to UCI website data donated at 1988. Its primarily used for classification machine learning tasks.

Dataset have total of 303 observations (rows) and 14 columns (features). Below is list of complete feature attributes:

1. age: Age in years (max:77, min:29)
2. sex: Sex (1:Male, 0:Female)
3. cp: Chest pain type
  - (a) Value 1: typical angina
  - (b) Value 2: atypical angina
  - (c) Value 3: non-anginal pain
  - (d) Value 4: asymptomatic
4. trestbps: Resting blood pressure (in mm Hg on admission to the hospital)
5. chol: Serum cholestoral in mg/dl
6. fbs: Fasting blood sugar  $> 120$  mg/dl (1:true, 0:false)
7. restecg: Resting electrocardiographic results
  - (a) Value 0: normal
  - (b) Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of  $> 0.05$  mV)
  - (c) Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
8. thalach: Maximum heart rate achieved
9. exang: Exercise induced angina (1:yes, 0:no)
10. oldpeak: ST depression induced by exercise relative to rest
11. slope: The slope of the peak exercise ST segment

- (a) Value 1: upsloping
  - (b) Value 2: flat
  - (c) Value 3: downsloping
12. ca: number of major vessels (0-3) colored by flourosopy
  13. thal: 3:normal, 6:fixed defect, 7:reversable defect
  14. target: Outcome (1:Heart disease, 0:Normal)