

STT 811

In-Class Assignment 6

This problem will use the Heart dataset. The target is the AHD field.

1. Some of the fields are not clear. Find a website which describes the fields and read their description.
2. Determine whether the AHD target field is balanced or not.
 - a. The field is relatively balanced (55% No, 45% Yes)
3. Ignoring the X field (just an id) determine which fields are truly numeric or categorical.
 - a. Numeric: Age, RestBP, Chol, MaxHR, Oldpeak,
 - b. Categorical: Sex, ChestPain, Fbs, RestECG, ExAng, Slope, Ca, Thal, AHD
4. Do some exploratory data analysis on a few of the variables:
 - a. Age

```
summary(heart$Age)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 29.00  48.00   56.00   54.44  61.00   77.00
```

- b. ChestPain

```
> table(heart$ChestPain)
asymptomatic  nonanginal  nontypical    typical
           144           86           50           23
```

- c. OldPeak

```
summary(heart$Oldpeak)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.00  0.00   0.80   1.04  1.60   6.20
```

- d. Thal

```
table(heart$Thal)
      fixed      normal reversable
        18       166       117
```