

homework1 packege

Jiayi Lian

August 31, 2019

Homework1

Jiayi Lian

Aug 31,2019

Problem 2

Part A

- 1 Data analysis projects
- 2 Raw data manipulation
- 3 Database

Firstly, I would like to go through some data analysis projects to practice application of ML methods, such as xgboost, nnw, lgb, etc. Secondly, I am willing to work on raw data sets, which have flaws(like: Missing data, type error, measure error) and are not manipulated to be neat and formal. It is good to learn how to make raw data to be informative. In addition, knowledge of database is helpful when handling data.

Part B

Noraml distribution:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad (1)$$

T distribution:

$$f(x|\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{\sqrt{\nu\pi}} \frac{1}{(1 + (\frac{x^2}{\nu}))^{(\frac{\nu+1}{2})}} \quad (2)$$

Cauchy distribution:

$$f(x|\theta, \sigma) = \frac{1}{\pi\sigma} \frac{1}{1 + (\frac{x-\theta}{\sigma})^2} \quad (3)$$

Problem 3

1 Do data analysis step by step. Rules: Data manipulation should be standard. Store raw data behind plots. Set an increasing layers of details and allow all details to be discovered. 2 Producing results. Rules: Explain how every result is produced and give each result an texture(Interpretation). Store intermediate results in a standard format if possible. 3 Version records and control. Rules: Record all versions of all external programs. Using a specific version of script. Note random seeds if necessary. 4 Give public access to data, results, versions, etc.

Problem 4

```
#See data frams of chickwts
str(chickwts)
```

```
## 'data.frame': 71 obs. of 2 variables:
## $ weight: num 179 160 136 227 217 168 108 124 143 140 ...
## $ feed : Factor w/ 6 levels "casein","horsebean",...: 2 2 2 2 2 2 2 2 2 2 ...
```

```
#Choose weight as Y id as X to produce the scatter plot
```

```
plot(x = seq(1:length(chickwts$weight)), y=chickwts$weight, type="p", ylab = "Weight", xlab = "id", main = "Chicken's weight")
```

