BT1101 Introduction to Business Analytics

Hands-on Tutorial 01

Basic Information of Hands-on Tutorial 12

Tutor : He Yuan (Henry) year-3 Business Analytics & Applied Maths

Time : @7pm on Tuesday

Weekly Routine: ① Review of Concepts : Handwritten notes, intuition, mathematical view 30-40 min

② Tutorial Q1 : practice with partners 15 min - 20 min

explain and work with me (needs participation!) 15 min - 20 min

③ Tutorial Q2 : analyze the problem (needs participation!) 15 min - 20 min

work on the problem together if time allows (needs participation!) TBD

Objectives: ① Encourage you to attempt the tutorial questions!

② Deepen your understanding of these Statistical and Data Mining Concepts!

③ Make friends here!

④ Ace in the exam and get A/A+!

Rules : ① Actively Participate! Interact with others and me! I'm not allowed to do the question alone!

② Optional to attend but highly recommended! Attendance will be taken.

③ Zoom will be recorded but will not be published (Q2 is excluded)

④ Handwritten notes and Codes will be uploaded to my github: @Henry-Yuan-He

⑤ Questions after class → book an ad-hoc coaching session / office hours with profs

Fromework of BT1101:

Basic Statistics + Basic Data Science Techniques

hugely broad!! 1 topic per lecture (Broad and a little bit deep)

essentially important foundation!

Ice Breaking Introduce yourself

Basics of R Programming

Data Structure in R:

1. Vector:
   - element type consistent!
   - How to name each element?
   - extract ele. by index / name / boolean
   - addition and sum
   - Concat two vectors

2. matrix:
   - initial a matrix
   - How to name the row / column
   - rowSums & colSums
   - Concat: cbind, rbind
   - multiplicity: * and %*%

3. factor: special vector
   - levels(···), nlevels(···)
   - Change original levels' names
   - use summary( )

   * Categorical variables:
     - nominal ~ : elephant, monkey, pig···
     - ordinal ~ : low, medium, high.

4. list. similar to vector
   - Allows different type elements!
   - How to name?
   - extract the element [[ ]] (index or name) or $
   - Append new element:
     - append(list, new element, position)
     - → return a new list.

5. data frame: ☆ important!

Dataset: $\{(\vec{x_t}, y_t)\}_{t=1}^{n}$ Feature Matrix: $X = [\overrightarrow{\text{client name}}\ \overrightarrow{\text{phone}}\ \overrightarrow{\text{Address}}\ \cdots]$

Attributes / Features / columns

observations / samples	Client Name	Phone	Address	.....	label y
1	xx	x	xxx	.....	Good
2	xxx	x	xxx	.....	Bad
3	xxxx	xx	null	.....	Bad
⋮					
n	xxx	null	x	....	Good

understand the structure:
- head(···, 3), str(···)
- tail(···, 5), View(···)

extract the claumn (with condition):
- df $···, [, ···]
- df %>% filter(condition)
- subset(df, condition)