BT5152 Tutorial 1

AY 2018/19, Semester 1, Week 3

Lu Wei

About Me

- TA for tutorial 1 - 5
- 7 years of industry experience as a software engineer
- Bachelor of Computing (Information Systems)
- Current Part-time CS Masters Program
- Email: weilu@comp.nus.edu.sg
- Feel free to interrupt me any time during tutorial if you have a question
- Post questions on forum or catch me before/after class

Dev Environment Setup

Options:

- Mac or Windows: RStudio
- Mac: brew install r (if you are comfortable with unix command line)

Key Concepts Revision

- Training vs Test Dataset
- Accuracy
- Categorical Variable Encoding
- Numerical Variable Scaling
- Naive Bayes
- KNN
- Decision Tree
- Decision Tree Pre-Pruning vs Post-Pruning

Tutorial Exercises: Swirl & R Basics

Open RStudio, in Console, type:

```
install.packages("swirl")
library(swirl)
install_course("A_(very)_short_introduction_to_R")
swirl()
```

Follow instructions. When asked to choose a lesson, start with `2: Module 1`

You may finish Module 2 and 3 at home for your own practice and knowledge.

https://swirlstats.com/students.html

Tutorial Exercises:
Train/Test Split, Classification Models

RStudio > Console:

```
install.packages("swirl")
library(swirl)
install_course_github('weilu', 'BT5152', multi=TRUE)
swirl()
```

1: Train Test Split

2: KNN and Naive Bayes

3: Decision Trees

Assignment 1

- Upload 2 files (*.R and *.pdf) into IVLE BT5152 workbin under the folder: Student Submission > A1
- Assume dataset files are in the same directory as the R script, i.e. `train_data <- read.csv("loan_train.csv", stringsAsFactors = TRUE)`
- Make sure your R file is runnable and has all the dependency packages imported e.g. `library(C50)`
- I will fail any submission with an R file that's not executable
- You may revise and submit as many times before deadline. Make sure to remove any old version that you don't wish to be graded.
- Plagiarism: 0 grade + reported to the department

Useful References

- dplyr cheatsheet for data manipulation: https://www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf
- More swirl courses: https://github.com/swirldev/swirl_courses