DATS 6450 - 11, Fall 2017, Homework 2
Due Wednesday Sep 20, 11:59 PM

Material Covered

- The R language

Submission:

- Submit a pdf file named `YourFirstName_YourLastName.pdf` and a R file named `Solution.R` to blackboard folder `/Homework_2/`.
- We expect you to follow a reasonable programming style. While we do not mandate a specific style, we require that your code to be neat, clear, **documented/commented** and above all consistent. **Marks will be deducted if these are not followed**.

Question 1: Programming in R (90 Points)

Problem

- Apply Simple Linear Regression on UCI Bike sharing dataset, publicly available at: `https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset`
- A very nice discussion of Simple Linear Regression can be seen in Chapter 3.1 of book *An Introduction to Statistical Learning* (ISL), publicly available at: `http://www-bcf.usc.edu/~gareth/ISL/`
- The goal for this problem is to learn the following linear model:

$$cnt = \beta_1 * temp + \beta_0,$$

  where *cnt* is the count of total rental bikes and *temp* the temperature (the description of the dataset can be seen on the website above)
- Specifically, you should calculate the coefficients, β_1 and β_0, using the *least squares coefficient estimates* (defined by eq.(3.4) on page 62 of ISL)

Deliverables

- Implement the following four functions in `Solution.R`:
  - `loadData`: load dataset and return the data of *cnt* and *temp*
  - `train`: train the model on the training set (`hour.csv`) and return the coefficients, β_1 and β_0
  - `test`: test the model on the testing set (`day.csv`) and return the *Residual Sum of Squares* (RSS, defined on page 62 of ISL)
  - `plotDataModel`: plot the data of *cnt* and *temp*, the linear model, and save the figures to `trainingResultFig.pdf` and `testingResultFig.pdf`
- The `Driver.R` and examples of the output figures are provided in blackboard folder `/Homework_2/`

Question 2: Discussion (10 Points)

- Based on the figure obtained on the training set, `trainingResultFig.pdf`, what conclusion can you make?
- As shown in the figure obtained on the testing set, `testingResultFig.pdf`, our model does not fit the testing data well. What could be the reason?