Assignment 0

CS 750/850 Machine Learning

- **Due**: Monday 1/27 at 11:59PM
- **Submisssion**: Turn in as a **PDF** and the **source code** (R,Rmd,py,ipynb) on MyCourses
- **Questions**: Piazza and Office hours: *Marek*: Wed 1:30-3:00pm, Soheil: Mon 2-4pm, Xihong: Thu 1:30-3:30pm
- **Extra credit**: Especially good questions or helpful answers on Piazza regarding the assignment earn up to 5 points extra credit towards the assignment grade.

Problem 1 [33%]

What are the advantages and disadvantages of very flexible (vs less flexible) approach for regression or classification?

1. When would be a more flexible approach preferable?
2. What about a less-flexible approach?

Problem 2 [33%]

Install and learn to use R (https://www.r-project.org/) or Python, read the labs in Chapter 2 of the textbook. We recommend that you use R Notebooks of RStudio to typeset homeworks. Jupyter is a comparable tool for Python. Use Python or another tool (like MATLAB or Julia) if you have some experience and you will not need help from the TA/instructor. Then:

1. Download the advertising dataset (`Advertising.csv`) from http://www-bcf.usc.edu/~gareth/ISL/data.html and load it into R/Python (use function `read.csv()` in R or Pandas in Python)
2. What are the minimum, maximum, and mean value of each feature? (in R use function `summary()` and or `range()`)
3. Produce a scatterplot matrix of all variables (in R use function `pairs()`)
4. Produce a histogram of TV advertising (in R use function `hist()`)

Problem 3 [34%]

Describe some real-life applications for machine learning.

1. Describe one real-life application in which *classification* combined with *prediction* may be useful. Describe the response and predictors.
2. Describe one real-life application in which *classification* combined with *inference* may be useful. Describe the response and predictors.
3. Describe one real-life application in which *regression* combined with *prediction* may be useful. Describe the response and predictors.
4. Describe one real-life application in which *regression* combined with *inference* may be useful. Describe the response and predictors.

Optional Problem O3 [39%]

This problem can be substituted for Problem 3 above, for 5 points extra credit. At most one of the problems 3 and O3 will be considered.

Read sections 1.2, 1.2.1, 1.2.2 in [Bishop, C. M. (2006). Pattern Recognition and Machine Learning] and solve *Exercise 1.5*.

Hints

1. An easy way to launch help for any function in R, such as `summary`, is to execute: `> ?summary`
2. See http://rmarkdown.rstudio.com/pdf_document_format.html for how to generate a PDF from an R notebook in R-studio. You will also need to install LaTeXwhich you can get from https://www.latex-project.org/get/
3. For more advanced (and prettier?) plotting capabilities, see the package `ggplot`: http://ggplot2.tidyverse.org/ and https://github.com/rstudio/cheatsheets/raw/master/data-visualization-2.1.pdf
4. If you think you may struggle with R, consider signing up for MATH 759, a 1-credit online introduction to R.