Statistical Learning for Engineers (EN.530.641)
Homework 3

Jin Seob Kim, Ph.D.

Out: 09/23/2022
due: 09/30/2022 by midnight EST

This is exclusively used for Fall 2022 EN.530.641 SLE students, and is not to be posted, shared, or otherwise distributed.

1. Let $A \in \mathbb{R}^{p\times p}$ and $\mathbf{x} \in \mathbb{R}^p$.

(a) Calculate

$$\frac{\partial}{\partial \mathbf{x}}\left(\mathbf{x}^T A\mathbf{x}\right)$$

where A is not symmetric.

(b) Then when A is symmetric, calculate

$$\frac{\partial}{\partial \mathbf{x}}\left(\mathbf{x}^T A\mathbf{x}\right)$$

to show what we had in the class.

2. Generate Figures 2.3 and 2.2 in the textbook (ESL) using k-nearest-neighbor (k-NN) classification. The data in csv format will be on the canvas.

(a) Write your own python code to implement k-NN and plot those figures.

(b) Use Scikit-Learn library to plot those figures. For example, you will need `KNeighborsClassifier` in `neighbors` in `sklearn`.

3. In this problem, you will generate your own data randomly. You have only 1-dimensional input, X (output Y is real, of course).

(a) Generate 200 uniformly distributed random numbers between 0 and 3 as x_i. Hint: look for `numpy.random.rand`). Then the corresponding y_i is generated as $Y = 4 + 3X + \epsilon$, where $\epsilon \sim \mathcal{N}(0, 1)$. For this random error term, look for `numpy.random.randn`. Use this data set for the next two parts.

(b) Write your own Python code to perform linear regression. Show your results (values of $\widehat{\boldsymbol{\beta}}$), and plot the data and your result (line) together. Also

(c) Use Linear Regression library in `sklearn` to perform linear regression. Compare your results with the previous part.

Submission Guideline

- For analytic parts (e.g., problems 1, plots of problem 2, and the results of problem 3 (b) and (c)), submit your homework answers in a single pdf format, including plots, to “HW3_analytical” on the gradescope.
- No more than two (2) homework problems may be on the same page. In other words, for each problem your answers should be on a separate set of pages. Then when submitting, you should assign the pages to each problem on Gradescope.
- *Show your work.*
- Submit all your python codes in a single .zip file that contains codes for each problem (name them by including the problem number). Name your single zip file submission as “YourName_HW3.zip”. For example, “JinSeobKim_HW3.zip” for a single zip file. Submission will be done through “HW3_computational” on the gradescope.
- Just in case you have related separate files, please make sure to include *all the necessary files*. If TAs try to run your function and it does not run, then your submission will have a significant points deduction.
- Make as much comments as possible so that the TAs can easily read your codes.