STAT 534 Homework 1
Due April 15, 2019
©2019 Marina Meilă
mmp@stat.washington.edu

Please read the **Assignments** *page on the course web site for instructions on how to submit your homework.*

Problem 1 – Basic statistics – NOT GRADED

a. Write a program that reads $n \leq 100$ floating point numbers from a file and computes their mean, standard deviation. Store the numbers in a `numpy` array. Write your own code for mean and standard deviation, using `numpy` array functions.

The file has the following format: the first line contains the value of n. On the next line start the floating point numbers, each on a new line. See also the example file `hw1-pb1.dat` (that you can download from the web page).

b. Run the program on the file `hw1-pb1.dat` and print out the results.

The program output should be informative to a human, e.g. `The data in hw1-pb1.dat has mean =...and stdev =....` *Edit the template given for Problem 4 to read data from a file.*

Problem 2 – Finding a number in a list – NOT GRADED

a. Write a program that:

1. Reads $n \leq 100$ floating point numbers from a file and stores them in memory; you can choose what data structure to use for storing the numbers.

2. asks you to enter another (floating point) number x

3. Finds if the number x is present in the stored set of numbers or not

and prints the number and the answer (e.g. "found!" or "not found!").

4. repeats from 2

The file has the same format as in Problem 1.

b. Run on the file `hw1-pb2.dat` with input data 0, 1.1, 2.2, 3.3 and print out the program output. *Edit the template given for Problem 4 to read data from a file.*

c. Explain in one sentence why you chose the data structure in 1,a.

Problem 3 – Computing the value of a polynomial – NOT GRADED

Write a program that:

1. reads a set of floating point numbers a_0, a_1, a_2, ... a_n, $n \leq 100$ from file `hw1-pb3-poly.dat`

2. asks you to enter another (floating point) number x

3. computes the value of the polynomial $P(x) = a_0 + a_1 x + a_2 x^2 + \ldots a_n x^n$ and prints it out

4. repeats from 2

Can you find a solution that uses only multiplication, not power?

b. Run the program for $x = 0, 1, 3.14159, 1001$ and print out the program output.

Problem 4 – Lists, Arrays and Strings in python

Write a python program that does the following.

a. Reads strings from a file (1 string per line). The file format is the same as in Problem 1. The strings are stored in a list `l10`. Let `n` be the number of

strings. In the homework, these strings will be in `hw1-statisticiansA-M.txt`, containing names of statisticians extracted from Wikipedia.

b. Truncates all strings to `k=3` characters. You can truncate them in place or by copying.

c. Now form a string `longstring` of all `n` names, by appending strings one by one to `longstring`.

d. Next, form another string `longlongstring`, of length `t*n*k`, that repeats step c. `t=100` times, i.e. keeps iterating over the list `t` times and appending. For this step, you should write pseudocode or give code snippet of your implementation.

e. Create an `numpy` array of appropriate size, and redo step d. by copying the strings into the array, in sequence. Write this as efficiently as you can. You are not expected to do something optimal here, but just to do something that is faster than d, and to know why this is so.

Describe how you implemented this step. OK to include a short snippet of relevant code in your homework, or pseudocode. For maximum credit be as clear and succint as possible.

f. Time these operations and output the run times of steps d, e.

g. `print( longstring )`
`print( longlongstring )`
Do not change these statements because this output will be graded by running your code.

What to submit: In the homework .pdf, answers to questions **d, e, f.** *Your python code in a file named* `hw1-pb5.py`. *It is important to use the exat file names for the pyton program, for* `statisticiansA-M.txt` *and to not rpint anything with thestrings, because your code will be run by a script.*