CS 5644: Assignment 1

Date Assigned: Sep 13, 2017

Date Due: Sep 27, 2017, midnight ET

**(Python)**

1. (15 points) Define a function is\_palindrome() in Python that recognizes palindromes (i.e. words that look the same written backwards). For example, is\_palindrome("radar") should return True. is\_palindrome("CS5644") should return False.
2. (15 points)In cryptography, a Caesar cipher is a very simple encryption technique in which each letter in the plain text is replaced by a letter some fixed number of positions down the alphabet. For example, with a shift of 3, A would be replaced by D, B would become E, and so on. The method is named after Julius Caesar, who used it to communicate with his generals. ROT-13 ("rotate by 13 places") is a widely used example of a Caesar cipher where the shift is 13. In Python, the key for ROT-13 may be represented by means of the following dictionary:

key = {'a':'n', 'b':'o', 'c':'p', 'd':'q', 'e':'r', 'f':'s', 'g':'t', 'h':'u',

'i':'v', 'j':'w', 'k':'x', 'l':'y', 'm':'z', 'n':'a', 'o':'b', 'p':'c',

'q':'d', 'r':'e', 's':'f', 't':'g', 'u':'h', 'v':'i', 'w':'j', 'x':'k',

'y':'l', 'z':'m', 'A':'N', 'B':'O', 'C':'P', 'D':'Q', 'E':'R', 'F':'S',

'G':'T', 'H':'U', 'I':'V', 'J':'W', 'K':'X', 'L':'Y', 'M':'Z', 'N':'A',

'O':'B', 'P':'C', 'Q':'D', 'R':'E', 'S':'F', 'T':'G', 'U':'H', 'V':'I',

'W':'J', 'X':'K', 'Y':'L', 'Z':'M'}

Your task in this exercise is to implement an encoder/decoder of ROT-13. Write one function/method to encode messages and write another to decode them. Once you're done, you will be able to read the following secret message:

Pnrfne pvcure? V zhpu cersre Pnrfne fnynq!

**(Data exploration)**

The following two questions involve the breast cancer dataset from the UCI machine learning repository. This dataset contains 700 instances, each with 8 features. The class information indicates whether the instance is benign or malignant. You will find all the necessary data in data.csv. A jupyter notebook is also provided along with the assignment that has a *helper* function to read data from the file into a NumPy array.

1. (15 points) Create a stacked histogram showing the clump thickness for malignant and benign tumors, with the malignant class colored red and the benign class colored blue.
2. (15 points) Create a scatter plot to visualize the relationship between clump thickness and single epithelial cell size, and color each instance red or blue based on whether it is malignant or benign, respectively.

(SQL)

1. (10 points) From the sample SQL query problems handout, solve question 9, i.e., find the name of the person who received the highest grade in the “Big Data” course. You should NOT use any SQL concept that was not covered in the lecture, e.g., you should not use operators like MAX, MIN, COUNT, subqueries, correlated subqueries, or any other fancy feature of SQL. This question is intended to encourage you to think in plain simple relational concepts.
2. (10 points) Solve question 10 from the sample SQL query problems handout. Same conditions as previous question apply.
3. (10 points) From the sample SQL query problems handout, solve question 11, i.e., find the name of people who received a grade of 3.5 or better in every course (given in the Courses table). We do not know how many courses are in the Courses table. If the table contains 3 courses, then we are interested in students who have received a grade of 3.5 or more in each of these 3 courses. If it contains 10 courses, then we are interested in those who have excelled in all these 10 courses. (You get the idea.) Once again, you do NOT need to use (and should not) any fancy SQL feature that was not covered in the lecture. Solve it using clean, simple, relational concepts.
4. (10 points) Solve question 12 from the sample SQL query problems handout.

**What to submit:**

One single Python notebook file containing answers to questions 1-4 as Python code, and answers to questions 5-8 as comments in the notebook. Thus, the SQL queries will be shown as comments in the notebook. The assignment will be graded by executing your code for questions 1-4 and by manual inspection of solutions to questions 5 and 6.

**Contents in the assignment package:**

1. A word document (this document) describing the assignment – questions and submission instructions
2. A data.csv file, which contains the data required for questions 3 & 4
3. A README.txt that describes the details of the data (Again for questions 3&4)
4. assignment1.ipnyb, a Jupyter notebook (that has a helper function to read the data) for completing the assignment, for questions 3 &4.