PATTERN RECOGNITION -- Spring 2019

Assignment 3: Clustering

DUE: Before 12midnight on 15 Apr 2019 (Monday)

INSTRUCTIONS:

i. Please do the assignment in Python.
ii. You need to submit pdf files to the TAs. One file should contain your answers, results and analysis. A separate file should contain code you have written and its sample output.
iii. At the top-right of the first page of your submission, include the assignment number, your name and roll number.
iv. *IMPORTANT:* Make sure that the assignment that you submit is your own work. *Do not directly copy any part from any source* including your friends, seniors or the internet.
v. Your grade will depend on the correctness of answers and output. In addition, due consideration will be given to the clarity and details of your answers and the legibility and structure of your code.

Preamble:

The aim of this assignment is to experiment with *clustering* techniques we learned in the class on real world problems.

(1) Apply K-means clustering on the following two datasets.
    a. Write your own code for K-means clustering.
    b. IRIS dataset (omit the class labels and perform clustering).
    c. One dataset of your choice from the UCI Machine Learning Repository
(2) There are two kinds of Cluster validation measures – Internal Measures and External Measures. Describe any two cluster validation measures of each kind. Apply these measures on your datasets as well as compute the Confusion Matrix. Present an analysis and discussion of your results.
(3) Use Kernel K-means from any python package (for example, *scikit-learn*) on the dataset you chose in (1)(c) and compare both the results.

As part of the submission include the code for each of the algorithms along with a small report that explains the algorithms, implementation details, the results and their analysis.