Alexandria University
Faculty of Engineering
Computer and Systems Engineering Dept.

Prof. Dr. Mohamed Abougabal
Prof Dr. Nagwa M. El-Makky
Dr. Marwan Torki
Special Topics in Information Systems

Assignment 1
Data Exploration

Objectives

This assignment aims to make students familiar with python programming and to begin the exploration process of a dataset.

Download and Read Data

In your python script use the ”datasets.load” to load the iris dataset. Where X = iris.data and y = iris.target

Cosine Similarity

The **Cosine similarity** is a measure of **similarity** between two non-zero vectors of an inner product space that measures the **cosine** of the angle between them

$$(X, Y) = < X, Y > /(||X|| * ||Y||)$$

Compute the cosine similarity matrix for the Iris data set, and comment on the resultant similarity matrix. Also use ”imshow” to visualize the resultant similarity matrix.

Visualization

- Plot the X data for each class alone.
- Plot the histogram for each class.
- Use scatter plot to plot every 2 attributes together.
- Use 3D scatter plot to plot every 3 attributes together

Notes

1. Investigate the Iris data set and explain the iris dictionary in your report.
2. Deliver a report having all the assigned visualizations and comment on each image.
3. Copied assignments will be penalized; so not delivering the assignment would be much better.
4. You should write your code in python and you may use scikit-learn, numpy, scipy or pyplot whenever needed.
5. You should work in groups of 2.
6. Assigned: 2/10/2018 Due: 9/10/2018

Good Luck