Math 285 HW3, Fall 2015.

Instructions:

- Due date: **Tuesday, November 10, in class**.
- Please type your homework in Word or LaTex.
- All the data needed by this homework can be found on the course website: `http://www.math.sjsu.edu/~gchen/math285.html`.
- For each programming question, submit both the Matlab script and the output (any numerics and/or figures).

Answer the following questions.

(1) Consider a weighted graph with weights

$$\mathbf{W} = \begin{pmatrix} 0 & .15 & .15 & .3 & 0 \\ .15 & .0 & .85 & 0 & 0 \\ .15 & .85 & 0 & 0 & 0 \\ .3 & 0 & 0 & 0 & .9 \\ 0 & 0 & 0 & .9 & 0 \end{pmatrix}$$

Do the following:
- Draw the graph
- Find the degrees of all vertices
- Consider the two possible partitions $V = \{v_1, v_2, v_3\} \cup \{v_4, v_5\}$ and $V = \{v_1, v_4, v_5\} \cup \{v_2, v_3\}$. Which one has a smaller Ncut? Which one has a smaller ratio cut?

(2) We showed in class that for any bipartition of a graph $V = A \cup B$, the Normalized Cut (NCut) criterion can be expressed as a ratio of the following form

$$\text{NCut}(A, B) = \frac{\mathbf{x}^T \mathbf{L} \mathbf{x}}{\mathbf{x}^T \mathbf{D} \mathbf{x}},$$

where $\mathbf{x}$ is a vector dependent on the partition A, B:

$$x_i = \begin{cases} \sqrt{\frac{\text{vol}(B)}{\text{vol}(A)}}, & i \in A \\ -\sqrt{\frac{\text{vol}(A)}{\text{vol}(B)}}, & i \in B \end{cases}$$

In this problem you are asked to show that the above equation still holds true if we modify the definition of $\mathbf{x}$ to

$$x_i = \begin{cases} \frac{1}{\text{vol}(A)}, & i \in A \\ -\frac{1}{\text{vol}(B)}, & i \in B \end{cases}$$

(3) Implement in Matlab the multiway NCut algorithm on your own to cluster the data in *fakeface.mat*.

(4) This question concerns the RatioCut criterion:

$$\text{RatioCut}(A, B) = \text{cut}(A, B) \left(\frac{1}{|A|} + \frac{1}{|B|} \right)$$

It can be shown that this criterion leads to the following problem (after a similar relaxation used by the normalized cut):

$$\min_{\substack{\mathbf{x}\in\mathbb{R}^n:\\ \mathbf{x}^T\mathbf{1}=0}} \frac{\mathbf{x}^T\mathbf{L}\mathbf{x}}{\mathbf{x}^T\mathbf{x}}.$$

A minimizer of the above relaxed problem is given by the second eigenvector $\mathbf{x}^*$ of the unnormalized graph Laplacian $\mathbf{L}$:

$$\mathbf{L}\mathbf{x}^* = \lambda_2\mathbf{x}^*.$$

Thus, one may instead use this eigenvector in the same way for clustering (i.e., by looking at the signs of the coordinates of $\mathbf{x}^*$). You are asked to implement also this version of graph cut (let's call it unnormalized spectral clustering) and compare it with the Ncut algorithm using the data *twogaussians_1L1S.mat*. Which algorithm performs better and what is the reason?

(5) This problem concerns how to compare several values of the scale parameter σ needed by spectral clustering for constructing a weighted graph:

$$w_{ij} = e^{-\frac{\|\mathbf{x}_i-\mathbf{x}_j\|^2}{2\sigma^2}}, \quad \forall\, i \neq j.$$

The technique is to perform spectral clustering with each choice of σ and compare the corresponding total scatter in the eigenvector space (after the kmeans step); the σ that leads to the smallest total scatter in the eigenvector space will be considered optimal. Here, you are asked to verify this idea by applying the multiway NCut algorithm that you implemented in Question 3 to the data *threecircles.mat*, with various values of σ below:

$$.02,\ .04,\ .06,\ .08,\ .1,\ .12$$

Display all the clustering you obtained and also plot the total scatter of the data in the eigenvector space against σ. Does the plot predict correctly which value(s) of σ should be used?