Mathematics for Data Sciences **Sep 30th, 2011**

Homework 2. Diffusion Map

Instructor: Xiuyuan Cheng, Yuan Yao

1. (Eigenvalues of row-stochastic matrix) Let A_{ij} be an n-by-n row-stochastic matrix, i.e $A_{ij} \geq 0$ and $\sum_j A_{ij} = 1$.

   (a) A is *irreducible*, iff $\forall (i,j)$, $\exists t$ positive integer, s.t. $(A^t)_{ij} > 0$. Show that if A is *irreducible*, then the eigenvalue 1 is simple.

   (b) A is *primitive*, iff $\exists t$ positive integer, s.t. $\forall (i,j)$, $(A^t)_{ij} > 0$. Show that if A is *primitive*, then -1 is not an eigenvalue of A.

   (c) If $A_{ii} > 0$ $\forall i$, and A is irreducible, then A is *primitive*. (Hint: Write $A = \varepsilon I + (1?\varepsilon)B$ with $\varepsilon = \frac{1}{2}\min_{1\leq i\leq n} A_{ii}$ and B irreducible, and develop A^t.) So Markov chains induced by Gaussian kernels $W_{ij} = \exp(-\|x_i - x_j\|^2/t)$ are always primitive.

2. (Diffusion Map of a ring graph) Suppose the n-by-n affinity matrix

$$W = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 & 1 \\ 1 & 1 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & 1 & 1 \\ 1 & 0 & \cdots & 0 & 1 & 1 \end{bmatrix}$$

   and $A = D^{-1}W = \frac{1}{3}W$. Show that

$$(v_k)_j = e^{i\frac{2\pi k}{n}j}, \; j = 1, \cdots, n$$

   is a (complex-valued) eigenvector of A with eigenvalue

$$\lambda_k = \frac{1}{3}(1 + 2\cos(\frac{2\pi}{n}k)),$$

   $k = 0, 1, \cdots, K$, $K = \frac{n}{2}$ if n even and $\frac{n-1}{2}$ if n odd. What are the real-valued eigenvectors and what is the multiplicity of λ_k? Discuss the case of even and odd n. Show that using the first two eigenvectors except the all ones one (which corresponds to $\lambda_0 = 1$) to construct the Diffusion Map, it maps the n data points to a circle in $\mathbb{R}^2$.

3. (Convergence of $A^{(1)}$) For d-dimensional manifold M in $\mathbb{R}^p$ (suppose without boundary) and a probability density $p(x)$ on it, let

$$k_\epsilon(x,y) = h\left(\frac{|x-y|^2}{\epsilon}\right), \quad h(\xi) = e^{-\frac{\xi}{2}},$$

$$d_\epsilon(x) = \int_M k(x,y)p(y)dy,$$

and

$$k_\epsilon^{(1)}(x,y) = \frac{k_\epsilon(x,y)}{d_\epsilon(x)d_\epsilon(y)},$$

and then the operator

$$(A^{(1)}f)(x) = \frac{\int_M k_\epsilon^{(1)}(x,y)f(y)p(y)dy}{\int_M k_\epsilon^{(1)}(x,y)p(y)dy},$$

where $f \in C^\infty(M)$. Show that, up to multiplying a positive constant,

$$\lim_{\epsilon\to 0} \frac{1}{\epsilon}(A^{(1)} - I)f = \frac{1}{2}\triangle_M f,$$

where $\triangle_M$ is the Laplace-Beltrami operator.

4. (Diffusion Distance) Show that $d_t(i,j) = 0$ but $i \neq j$, if and only if node i and j have the same neighbors and proportional weights, *i.e.* $w_{ik} = \alpha w_{jk}$ for some $\alpha > 0$ and for all $k = 1, \ldots, n$.

5. (Untangle the knot) The trefoil knot curve has the following parametric equation (`http://en.wikipedia.org/wiki/Trefoil_knot`)

$$\begin{cases} x(t) &= (2+\cos(3t))\cos(2t) \\ y(t) &= (2+\cos(3t))\sin(2t) \ , \quad 0 \le t \le 2\pi, \\ z(t) &= \sin(3t) \end{cases}$$

which is a closed curve in $\mathbb{R}^3$. Randomly sample $n = 400$ data points along the curve, 'along' means the points roughly cover the whole curve, though not necessarily evenly distributed. (One possible way to do this is to randomly generate 400 points in $[0,1]$ first, and then map them to $\mathbb{R}^3$ using the parametric equation.) Use Gaussian kernel, that is

$$W_{ij} = e^{-\frac{|x_i - x_j|^2}{\epsilon}},$$

and build both $A^{(1/2)}$ (to approximate the Fokker-Plank operator) and $A^{(1)}$ (to approximate the Laplace-Beltrami operator) to construct the 2D diffusion map. Plot the two embedding and explain how and why they are different. (Hint: choose ϵ reasonably.)

6. (Order the faces) The UMIST face dataset face.mat can be found at course web in lecture 5. In Matlab, firstly

```
>> load face.mat;
```

then the Y(:,:,1:33) variable contains 33 pictures of a same person turing her face 90 degree. The pictures in Y(:,:,1:33) are randomly permuted from its correct order. The task is to construct a 1D embedding, who gives a value v_j to the j-th picture, $j = 1, \cdots, 33$, where the correct order of the pictures can be re-found from the order of v_j. That means, for example, use Matlab function *sort*

```
>> [ ~, ind ] = sort(v, 'ascend');
```

to get the index of sorting, then Y(:,:, ind) gives the sorted pictures. One can use the first nontrivial eigenvector to construct a Diffusion Map (using $A^{(0)}$ or $A^{(1)}$) or anything else. New methods are welcome and creative solutions win extra credits. You can also use more than one methods. Show the result and explain what you do. If you use other data set please explain the source of data.