Lecture 24

9.6.3 Relating graph embeddings and manifold embeddings

We now related optimal graph embeddings to optimal manifold embeddings. To do so we shall leverage the heat kernel. Let $f : \mathcal{M} \to \mathbb{R}$ be the initial heat distribution on a manifold $\mathcal{M}$, and let $u(x,t)$ be the heat distribution at time t so that $u(x,0) = f(x)$. Recall the heat equation is

$$\left(\frac{\partial}{\partial t} + \mathcal{L}\right) u = 0,$$

where $u(x,t)$ is a function of space $x \in \mathcal{M}$ (the manifold) and time $t \in [0,\infty)$, and where $\mathcal{L} = -\Delta$. Recall as well that the solution is given by

$$u(x,t) = \int_{\mathcal{M}} K(x,y;t) f(y)\, dV(y) \tag{66}$$

where $K(x,y;t)$ is the heat kernel. It thus follows that:

$$\mathcal{L}f(x) = \lim_{t_0 \to 0} \mathcal{L}u(x,t_0) = -\lim_{t_0 \to 0} \left(\frac{\partial}{\partial t} \int_{\mathcal{M}} K(x,y;t) f(y)\, dV(y)\right)_{t=t_0} \tag{67}$$

One can show that locally, the heat kernel even on a d-dimensional manifold looks like a Gaussian:

$$K(x,y;t) = (4\pi t)^{-d/2} e^{-\|x-y\|^2/4t} (\phi(x,y) + O(t)),$$

where $\phi(x,y)$ is a smooth function with $\phi(x,x) = 1$. Therefore,

$$K(x,y;t) \approx (4\pi t)^{-d/2} e^{-\|x-y\|^2/4t}, \quad t \text{ small.} \tag{68}$$

Notice that (66) implies that

$$f(x) = \lim_{t \to 0} \int_{\mathcal{M}} K(x,y;t) f(y)\, dV(y).$$

Let us now estimate the time derivative on the right hand side of (67). We have:

$$-\left(\frac{\partial}{\partial t} \int_{\mathcal{M}} K(x,y;t) f(y)\, dV(y)\right)_{t=t_0} \approx \ldots$$

$$\ldots \approx -\frac{1}{t} \left(\int_{\mathcal{M}} K(x,y;t_0+t) f(y)\, dV(y) - \int_{\mathcal{M}} K(x,y;t_0) f(y)\, dV(y)\right), \ t \text{ small.}$$

Thus taking the limit as $t_0 \to 0$ we get:

$$\begin{aligned}\mathcal{L}f(x) &= -\lim_{t_0 \to 0} \left(\frac{\partial}{\partial t} \int_{\mathcal{M}} K(x,y;t) f(y)\, dV(y) \right)_{t=t_0} \approx \dots \\ &\approx -\frac{1}{t} \left(\int_{\mathcal{M}} K(x,y;t) f(y)\, dV(y) - f(x) \right), \quad t \text{ small.} \\ &\approx \frac{1}{t} \left(f(x) - (4\pi t)^{-d/2} \int_{\mathcal{M}} e^{-\|x-y\|^2/4t} f(y)\, dV(y) \right), \quad t \text{ small.}\end{aligned}$$

Notice the normalization term $(4\pi t)^{-d/2}$ depends on the dimension d of the manifold. However, since $\mathcal{L}\mathbf{1} = 0$, we have:

$$0 \approx \frac{1}{t} \left(1 - (4\pi t)^{-d/2} \int_{\mathcal{M}} e^{-\|x-y\|^2/4t} f(y)\, dV(y) \right)$$

which implies

$$(4\pi t)^{d/2} \approx \int_{\mathcal{M}} e^{-\|x-y\|^2/4t}\, dV(y).$$

Thus if we set

$$q(x) = \int_{\mathcal{M}} e^{-\|x-y\|^2/4t}\, dV(y)$$

we get

$$\mathcal{L}f(x) \approx \frac{1}{t} \left(f(x) - \frac{1}{q(x)} \int_{\mathcal{M}} e^{-\|x-y\|^2/4t} f(y)\, dV(y) \right), \quad t \text{ small.} \tag{69}$$

Now suppose that $\mathcal{X} = \{x_1, \dots, x_n\} \subset \mathcal{M}$ are uniformly sampled from the manifold $\mathcal{M}$. Define:

$$q_n(x_i) = \sum_{j=1}^{n} e^{-\|x_i - x_j\|^2/4t}.$$

The right hand side of (69) can be approximated as:

$$\mathcal{L}f(x_i) \approx \frac{1}{t} \left(f(x_i) - \frac{1}{q_n(x_i)} \sum_{j=1}^{n} e^{-\|x_i - x_j\|^2/4t} f(x_j) \right)$$

Now let's use $\mathcal{X}$ as the vertices in a weighted graph $G = (\mathcal{X}, E, w)$. We take G to be a complete graph, with weighted edges:

$$w(i,j) = e^{-\|x_i - x_j\|^2/4t}. \tag{70}$$

Define the asymmetric normalized graph Laplacian as:

$$\overline{L} = I - D^{-1}A.$$

Notice that

$$D_{ii} = q_n(x_i).$$

Thus, if we define $\vec{f} \in \mathbb{R}^n$ as:

$$\vec{f}[i] = f(x_i),$$

it follows that

$$\mathcal{L}f(x_i) \approx \frac{1}{t}\overline{L}\vec{f}[i].$$

Thus the (asymmetric) normalized graph Laplacian is an approximation of the Laplace-Beltrami operator!

Notice we have the following relationship between the eigenvectors/eigenvalues of N and $\overline{L}$. Once again let φ_k be an eigenvector of N with eigenvalue γ_k,

$$(I - D^{-1/2}AD^{-1/2})\varphi_k = \gamma_k\varphi_k.$$

It follows that

$$\psi_k = D^{-1/2}\varphi_k$$

is an eigenvector of $\overline{L}$; indeed:

$$\begin{aligned}
\overline{L}\psi_k &= (I - D^{-1}A)D^{-1/2}\varphi_k \\
&= D^{-1/2}(D^{1/2} - D^{-1/2}A)D^{-1/2}\varphi_k \\
&= D^{-1/2}(I - D^{-1/2}AD^{-1/2})\varphi_k \\
&= D^{-1/2}\gamma_k\varphi_k \\
&= \gamma_k\psi_k.
\end{aligned}$$

Thus the Laplacian Eigenmaps algorithm computes eigenvectors which approximate the eigenfunctions of the Laplace-Beltrami operator, and uses them to embed the graph! In particular, the optimal graph embedding of the weighted graph G is, in a precise sense, the discrete analogue of the optimal embedding of the manifold $\mathcal{M}$ if the vertices of G are data points uniformly sampled from $\mathcal{M}$, and if the weights in the graph are given by (70).

Exercise 37 (optional). Using your Laplacian Eigenmaps code from the previous exercise (from the previous lecture), compute the Laplacian Eigenmaps embedding for the circles data set from Exercise 11. Use the weights

$$w(i,j) = e^{-\|x_i - x_j\|^2/4t}$$

for a small value of t (you will need to be somewhat careful here). Do you get a 2D manifold? Compare with your PCA result from Exercise 11.

References

[1] Bernhard Schölkopf and Alexander J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Adaptive Computation and Machine Learning. The MIT Press, 2002.

[2] Afonso S. Bandeira. Ten lectures and forty-two open problems in the mathematics of data science. MIT course *Topics in Mathematics of Data Science*, 2015.

[3] Jon Shlens. A tutorial on principal component analysis. arXiv:1404.1100, 2014.

[4] Karl Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine, Series 6*, 2(11):559–572, 1901.

[5] V. A. Marchenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, 72(114):507–536, 1967.

[6] J. Baik, G. Ben-Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(5):1643–1697, 2005.

[7] Debashis Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, pages 1617–1642, 2007.

[8] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer-Verlag New York, 2nd edition, 2009.

[9] Athanasios Tsanas and Angeliki Xifara. Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy and Buildings*, 49:560–567, 2012.

[10] Daniel A. Spielman. Spectral graph theory. *Yale Course Notes*, Fall, 2009.

[11] Jianzhong Wang. *Geometric Structure of High-Dimensional Data and Dimensionality Reduction*. Springer, 2012.

[12] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

[13] Ronald R. Coifman and Stéphane Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21:5–30, 2006.