On the Extension of Eigenvectors to New Datapoints

Christopher K. I. Williams
School of Informatics, University of Edinburgh
5 Forrest Hill, Edinburgh EH1 2QL, UK
c.k.i.williams@ed.ac.uk

November 18, 2006

Abstract

In this note we consider a $n \times n$ kernel matrix K_{nn} and the submatrix K_{mm} (for $m < n$) obtained by selecting m rows/columns of K_{nn}. We show that the Nyström approximation (Williams and Seeger, 2001) of the top m eigenvectors of K_{nn} is equivalent to that obtained from a variational argument based on Rayleigh's principle.

Consider a symmetric positive semidefinite kernel $k(\mathbf{x}, \mathbf{y})$ as used, for example, in support vector machines and Gaussian processes. Given a set of data points $\mathcal{D}_m = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$, we set up the matrix eigenproblem

$$K_{mm}\mathbf{u}_i^{(m)} = \lambda_i^{(m)}\mathbf{u}_i^{(m)}, \qquad i = 1, \ldots, m \tag{1}$$

where K_{mm} is a $m \times m$ matrix with entries $k(\mathbf{x}_i, \mathbf{x}_j)$, and $\mathbf{u}_i^{(m)}$ is the ith eigenvector of this matrix with eigenvalue $\lambda_i^{(m)}$. We assume the eigenvalues are ordered so that $\lambda_1^{(m)} \geq \lambda_2^{(m)} \cdots \geq \lambda_m^{(m)}$.

Now consider the case where we have a $n \times n$ kernel matrix, based on a dataset $\mathcal{D}_n = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, where the first m datapoints are the same as in $\mathcal{D}_m$, and the remaining $u = n - m$ are new, previously *unseen* datapoints. This matrix will have n eigenvectors and eigenvalues. We can extend the m eigenvalues/vectors of K_{mm} up to the full n dimensions using the Nyström approximation described in Williams and Seeger (2001) to give for $i = 1, \ldots, m$

$$\hat{\lambda}_i^{(n)} \stackrel{def}{=} \frac{n}{m}\lambda_i^{(m)}, \tag{2}$$

$$\hat{\mathbf{u}}_i^{(n)} \stackrel{def}{=} \sqrt{\frac{m}{n}}\frac{1}{\lambda_i^{(m)}}K_{nm}\mathbf{u}_i^{(m)} = \sqrt{\frac{m}{n}}K_{nm}K_{mm}^{-1}\mathbf{u}_i^{(m)}. \tag{3}$$

Note that the entries of $\hat{\mathbf{u}}_i^{(n)}$ for the points in $\mathcal{D}_m$ are just rescaled versions of $\mathbf{u}_i^{(m)}$.

We now consider a very different way to derive the same approximation. Let V_m be a $m \times q$ matrix. It is well known (Rayleigh's principle) that the solution of

$$\max_{V_m} \operatorname{tr}(V_m^T K_{mm} V_m) \text{ such that } V_m^T V_m = I_q \tag{4}$$

is that V_m is composed of the *top* q eigenvectors of K_{mm}, i.e. those corresponding to the q largest eigenvalues. The top q eigenvectors are also the solution of

$$\min_{V_m} \operatorname{tr}(V_m^T K_{mm}^{-1} V_m) \qquad \text{such that } V_m^T V_m = I_q; \tag{5}$$

notice here that max has been turned into min, and K_{mm} into K_{mm}^{-1}.

This variational formulation gives rise to an alternative way to extend an eigenvector to a new datapoint. Consider the $n \times n$ matrix K_{nn} and its inverse K_{nn}^{-1}. These are partitioned as

$$K_{nn} = \begin{pmatrix} K_{mm} & K_{mu} \\ K_{um} & K_{uu} \end{pmatrix}, \qquad K_{nn}^{-1} \stackrel{def}{=} \begin{pmatrix} \tilde{K}_{mm} & \tilde{K}_{mu} \\ \tilde{K}_{um} & \tilde{K}_{uu} \end{pmatrix}. \tag{6}$$

Similarly, we partition the $n \times q$ matrix V_n as

$$V_n = \begin{pmatrix} V_m \\ V_u \end{pmatrix}, \tag{7}$$

where V_u is $u \times q$. We now fix V_m to be the *top* q eigenvectors of K_{mm} and minimize $J = \mathrm{tr}(V_n^T K_{nn}^{-1} V_n)$ with respect to V_u. (We do not impose further constraints, but note that V_m is already orthogonal.) We obtain

$$J = \mathrm{tr}(V_m^T \tilde{K}_{mm} V_m) + \mathrm{tr}(V_m^T \tilde{K}_{mu} V_u) + \mathrm{tr}(V_u^T \tilde{K}_{um} V_m) + \mathrm{tr}(V_u^T \tilde{K}_{uu} V_u). \tag{8}$$

Minimizing this quadratic form by differentiating this wrt V_u gives

$$V_u = -\tilde{K}_{uu}^{-1} \tilde{K}_{um} V_m \tag{9}$$

Using the partitioned matrix inverse equations (see, e.g. Press et al. (1992, p. 77)) and particularly $\tilde{K}_{um} = -\tilde{K}_{uu} K_{um} K_{mm}^{-1}$ we obtain

$$V_u = K_{um} K_{mm}^{-1} V_m, \quad \text{or} \quad V_n = K_{nm} K_{mm}^{-1} V_m. \tag{10}$$

This is, of course, just the same solution that we obtained via the Nyström method in equation 3, up to scaling factors.

The variational method used above was inspired by the work of Ham et al. (2005) and Yang et al. (2006) on eigenmethods in nonlinear dimensionality reduction[1]. However, note that these authors were concerned with extending the *bottom* eigenvectors, i.e. those corresponding to the smallest eigenvalues. This leads to to a different minimzation problem (based on K_{nn} rather than K_{nn}^{-1}) and thus to a different solution.

Acknowledgements

I thank Miguel Carreira-Perpiñán for helpful discussions and for making available an advance copy of his paper on the Laplacian Eigenmaps Latent Variable Model.

References

Ham, J., Lee, D. D., and Saul, L. K. (2005). Semisupervised Alignement of Manifolds. In *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics (AISTATS 2005)*.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C*. Cambridge University Press, Second edition.

[1]There are some errors in the derivations of Yang et al. (2006) concerning missing trace operators, and a sign error in their eq. 12.

Williams, C. K. I. and Seeger, M. (2001). Using the Nyström Method to Speed Up Kernel Machines. In Leen, T. K., Diettrich, T. G., and Tresp, V., editors, *Advances in Neural Information Processing Systems 13*, pages 682–688. MIT Press.

Yang, X., Fu, H., Zha, H., and Barlow, J. (2006). Semi-Supervised Nonlinear Dimensionality Reduction. In *Proceedings of th 23rd International Conference on Machine Learning (ICML 2006).*