STA 414/2104

Statistical Methods for Machine Learning and Data Mining

Radford M. Neal, University of Toronto, 2012

Week 12

Kernel PCA

PCA on Basis Function Values

Rather than do PCA on the original vector of p values, x, we can do it on the values of m basis functions, $\phi(x) = [\phi_1(x), \ldots, \phi_m(x)]$.

If $m > p$, the basis function values will lie in a p-dimensional space embedded in an m-dimensional space. This would make modeling the data as a multivariate Gaussian be drastically incorrect, but PCA makes no distributional assumptions — it just finds the directions of maximum sample variance.

If m is big, we should of course use the trick that was presented earlier for when p is big — find the eigenvectors of the $n \times n$ matrix XX^T rather than the eigenvectors of the $p \times p$ matrix X^TX.

Note that even if the x values have been centred to have sample mean of zero, the $\phi(x)$ values will probably not be centred. So we'll have to subtract the sample mean for each basis function before finding eigenvectors.

Details of PCA on Basis Function Values

Let Φ be the $n \times m$ matrix of basis function values for the n observed items, so $\Phi_{ik} = \phi_k(x_i)$.

If we let $\mathbf{1}_n$ be a vector of n ones, we can get a row vector of sample means of the basis functions as $(1/n)\mathbf{1}_n^T\Phi$. The matrix of centred basis function values can then be written as

$$\widetilde{\Phi} \;\;=\;\; \Phi \;-\; \mathbf{1}_n[(1/n)\mathbf{1}_n^T\Phi] \;\;=\;\; [I_{n\times n} \;-\; \mathbf{1}_{n\times n}/n]\,\Phi$$

where $I_{n\times n}$ is the $n \times n$ identity matrix and $\mathbf{1}_{n\times n}$ is the $n \times n$ matrix of all ones.

We now find the eigenvectors of

$$\widetilde{\Phi}\widetilde{\Phi}^T \;\;=\;\; [I_{n\times n} \;-\; \mathbf{1}_{n\times n}/n]\,\Phi\Phi^T\,[I_{n\times n} \;-\; \mathbf{1}_{n\times n}/n]$$

If v is such an eigenvector, of length one, with eigenvalue λ, then $\widetilde{\Phi}^T v/\sqrt{\lambda}$ is an eigenvector of $\widetilde{\Phi}^T\widetilde{\Phi}$, also of length one, and with eigenvalue λ.

Projections of Basis Function Vectors for Test Points

What we're usually interested in are the projections of the basis function vectors for test points on the principal components.

Let $\widetilde{\Phi}^T v/\sqrt{\lambda}$ be a principal component direction, where v is an eigenvalue of $\widetilde{\Phi}\widetilde{\Phi}^T$ with eigenvalue λ. The projection in this direction of the centred basis function values for a point x_* is

$$\begin{aligned}
[\phi(x_*)^T - (1/n)\mathbf{1}_n^T\Phi]\,\widetilde{\Phi}^T v/\sqrt{\lambda} &= [\phi(x_*)^T - (1/n)\mathbf{1}_n^T\Phi]\,\Phi^T[I_{n\times n} - \mathbf{1}_{n\times n}/n]\,v/\sqrt{\lambda} \\
&= [\phi(x_*)^T\Phi^T - \mathbf{1}_n^T\Phi\Phi^T/n]\,[I_{n\times n} - \mathbf{1}_{n\times n}/n]\,v/\sqrt{\lambda}
\end{aligned}$$

Applying the Kernel Trick

All these operations involve $\phi(x)$ only via inner products. We can define

$$K(x, x') \;=\; \phi(x)^T \phi(x')$$

and then define the $n \times n$ matrix K by $K_{ij} = K(x_i, x_j)$. We then can compute

$$\widetilde{K} \;=\; \widetilde{\boldsymbol{\Phi}}\widetilde{\boldsymbol{\Phi}}^T \;=\; [I_{n\times n} - \mathbf{1}_{n\times n}/n]\, K\, [I_{n\times n} - \mathbf{1}_{n\times n}/n]$$

If the $n \times n$ matrix $\widetilde{K}$ has unit length eigenvectors $v_1, v_2, \ldots, v_n$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, then the projection of a data point x_* on the m'th principal component is

$$[k - \mathbf{1}_n^T K/n]\, [I_{n\times n} - \mathbf{1}_{n\times n}/n]\, v_m/\sqrt{\lambda_m}$$

where k is the vector of dimension n with $k_i = K(x_*, x_i)$.

Since ϕ no longer appears explicitly in these formulas, we can let the number of basis functions go to infinity, as long as we know how to compute $K(x, x')$.

Example of Kernel PCA

Kernel PCA for 2D data from two classes, using $K(x, x') = \exp(-||x - x'||^2)$.

Original data and pairs of projections on PC1, PC2, and PC3: