# Kernel Principal Components

Stephen Vardeman

Analytics Iowa LLC

ISU Statistics and IMSE

# PCs after transformation to a Euclidean space

Consider first the possibility of using a nonlinear function $\boldsymbol{\phi} : \Re^p \rightarrow \Re^M$ to map data vectors $\mathbf{x}$ to (a usually higher-dimensional) vector of features $\boldsymbol{\phi}(\mathbf{x})$. This creates a new $N \times M$ data/feature matrix

$$\boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\phi}'(\mathbf{x}_1) \\ \boldsymbol{\phi}'(\mathbf{x}_2) \\ \vdots \\ \boldsymbol{\phi}'(\mathbf{x}_N) \end{pmatrix}$$

with entries of $\boldsymbol{\Phi}$ belonging to $\Re$. After centering via

$$\widetilde{\boldsymbol{\Phi}} = \boldsymbol{\Phi} - \frac{1}{N}\mathbf{J}\boldsymbol{\Phi} = \left(\mathbf{I} - \frac{1}{N}\mathbf{J}\right)\boldsymbol{\Phi}$$

for $\mathbf{J}$ an $N \times N$ matrix of 1s, a SVD of $\widetilde{\boldsymbol{\Phi}}$, produces singular values and both sets of singular vectors for the new feature matrix.

# Centered features in a reproducing kernel space

Now suppose $\mathcal{K}$ is a kernel function and one maps data vectors $\mathbf{x}$ to elements $T(\mathbf{x}) = \mathcal{K}(\mathbf{x}, \cdot)$ in the abstract (function) feature space $\mathcal{A}$. One can think of finding "principal components" for the transformed training set in this feature space. First, the function

$$\overline{\mathcal{K}}(\cdot) \equiv \frac{1}{N}\sum_{i=1}^{N}\mathcal{K}(\mathbf{x}_i, \cdot)$$

is a well-defined linear combination of the images of the training set in $\mathcal{A}$ and therefore a sensible "center" of the transformed training set. The functions

$$\mathcal{K}(\mathbf{x}_i, \cdot) - \overline{\mathcal{K}}(\cdot)$$

are then sensible centered abstract feature values for the training set.

# Matrix of inner products of centered features

Corresponding to the matrix of inner products for a centered set of $N$ points in a Euclidean space is the $N \times N$ matrix of inner products of centered features in $\mathcal{A}$,

$$\mathbf{C} \equiv \left( \left\langle \mathcal{K}\left(\mathbf{x}_i, \cdot\right) - \overline{\mathcal{K}}\left(\cdot\right), \mathcal{K}\left(\mathbf{x}_j, \cdot\right) - \overline{\mathcal{K}}\left(\cdot\right) \right\rangle_{\mathcal{A}} \right)_{\substack{i=1,\dots,N \\ j=1,\dots,N}}$$

Then using the fact that $\left\langle \mathcal{K}\left(\mathbf{x}, \cdot\right), \mathcal{K}\left(\mathbf{z}, \cdot\right) \right\rangle_{\mathcal{A}} = \mathcal{K}\left(\mathbf{x}, \mathbf{z}\right)$ and the notation $\mathbf{K}$ for the Gram matrix with entries $\mathcal{K}\left(\mathbf{x}_i, \mathbf{x}_j\right)$, one gets the representation

$$\mathbf{C} = \mathbf{K} - \frac{1}{N}\mathbf{JK} - \frac{1}{N}\mathbf{KJ} + \frac{1}{N^2}\mathbf{JKJ}$$

Eigen analysis of $\mathbf{C}$ produces principal components ($N$ vectors of length $N$ of scores) for the training data expressed in the abstract feature space.

# Principal component “direction” functions

To realize the entries in these eigen vectors of kernel principal component scores as inner products of the $N$ functions $\mathcal{K}(\mathbf{x}_i, \cdot) - \overline{\mathcal{K}}(\cdot)$ with "principal component directions" in the abstract feature space, $\mathcal{A}$, one may return to Section 2.3.1 and begin with any orthonormal basis (coming, for example, from use of the Gram-Schmidt process) $\mathcal{E}_1(\cdot), \mathcal{E}_2(\cdot), \ldots, \mathcal{E}_N(\cdot)$ for the span of the $\mathcal{K}(\mathbf{x}_i, \cdot) - \overline{\mathcal{K}}(\cdot)$. Then the general inner product space argument beginning with an $N \times N$ matrix with entries $\left\langle \mathcal{K}(\mathbf{x}_i, \cdot) - \overline{\mathcal{K}}(\cdot), \mathcal{E}_j(\cdot) \right\rangle_{\mathcal{A}}$ produces $N$ basis functions $\mathcal{V}_1(\cdot), \mathcal{V}_2(\cdot), \ldots, \mathcal{V}_N(\cdot)$ whose $\mathcal{A}$ inner products with functions $\mathcal{K}(\mathbf{x}_i, \cdot) - \overline{\mathcal{K}}(\cdot)$ are (up to a sign for each $\mathcal{V}_j(\cdot)$) the entries of the eigen vectors of $\mathbf{C}$. In cases of small $p$ it may be of interest to examine the nature of these abstract principal component "direction" functions through some plotting.