CSE 881: How to Center a Kernel Matrix

Consider a data set $\mathcal{D}$ that contains N data points, $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$. Let $\Phi(\mathbf{x}_i)$ denote a feature representation of the data point $\mathbf{x}_i$. A kernel matrix (also known as Gram matrix) $\mathbf{K}$ can be computed between every pair of points in $\mathcal{D}$ by taking the dot product of their feature representation, i.e., for the $(i,j)-$th element of the matrix:

$$\mathbf{K}_{ij} = \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j), \tag{1}$$

where the superscript T denote the transpose operation. Note that $\Phi(\mathbf{x})$ is a column vector whose length could be infinite-dimensional.

A feature vector is said to be centered if we subtract it with the mean value of each column in the data matrix. Mathematically, the centering operation can be written as follows:

$$\Phi_c(\mathbf{x}_i) = \Phi(\mathbf{x}_i) - \frac{1}{N}\sum_{j=1}^{N} \Phi(\mathbf{x}_j).$$

To simplify the notation, let $\mathbf{\Phi} = [\Phi(\mathbf{x}_1)^T; \Phi(\mathbf{x}_2)^T; \cdots, \Phi(\mathbf{x}_N)^T]$ be the matrix representation of the feature vector for all the data points in $\mathcal{D}$. Using the centering operator, we can subtract the values in each column of $\mathbf{\Phi}$ with its corresponding column mean by applying the following matrix operation:

$$\mathbf{\Phi_c} = (\mathbf{I}_N - \frac{1}{N}\mathbf{1}_N \mathbf{1}_N^T)\mathbf{\Phi}, \tag{2}$$

where $\mathbf{\Phi_c}$ is the centered matrix, $\mathbf{I}_N$ is an $N \times N$ identity matrix, and $\mathbf{1}_N$ is an N-dimensional vector whose elements are all 1s.

Our goal is to derive an expression for the kernel matrix of the centered data in terms of the kernel matrix of its uncentered data. The uncentered kernel matrix is given by the formula in Equation (1). By using Equation (2), the kernel matrix of the centered data can be derived as follows:

$$\begin{aligned}
\mathbf{K}_c &= \mathbf{\Phi}_c \mathbf{\Phi}_c^T \\
&= (\mathbf{I}_N - \frac{1}{N}\mathbf{1}_N \mathbf{1}_N^T)\mathbf{\Phi}\left[(\mathbf{I}_N - \frac{1}{N}\mathbf{1}_N \mathbf{1}_N^T)\mathbf{\Phi}\right]^T \\
&= (\mathbf{I}_N - \frac{1}{N}\mathbf{1}_N \mathbf{1}_N^T)\mathbf{\Phi}\mathbf{\Phi}^T(\mathbf{I}_N - \frac{1}{N}\mathbf{1}_N \mathbf{1}_N^T)^T \\
&= (\mathbf{I}_N - \frac{1}{N}\mathbf{1}_N \mathbf{1}_N^T)\mathbf{K}(\mathbf{I}_N - \frac{1}{N}\mathbf{1}_N \mathbf{1}_N^T) \\
&= \mathbf{K} - \frac{1}{N}\mathbf{EK} - \frac{1}{N}\mathbf{KE} + \frac{1}{N^2}\mathbf{EKE},
\end{aligned} \tag{3}$$

where $\mathbf{E} = \mathbf{1}_N \mathbf{1}_N^T$ is an $N \times N$ matrix whose elements are all 1s. This is a nice result because it allows us to compute the centered kernel matrix by multiplying it with a known, finite-dimensional matrix $\mathbf{E}$, irrespective of the feature representation $\mathbf{\Phi}$.

In kernel PCA, you need to compute the eigenvalues and eigenvectors of a normalized and centered kernel matrix:

$$\frac{1}{N}\mathbf{K}_c\alpha = \lambda\alpha$$

You can use (3) on the left-hand side of the equation and derive the eigenvectors to find the kernel principal components.