# Kernel Methods

Homework 2

Luis Antonio Ortega Andrés
Antonio Coín Castro

March 4, 2021

## 1 TODO

Contestar preguntas en notebook (y eliminar este pdf).

Pasar .pys a pdf.

Añadir referencias (documentación sklearn, paper PCA?) en formato Chicago.

Aclarar que eliminamos las componentes no nulas en KPCA.

Añadir gráficas sobre cómo es KPCA según gamma->0 o gamma->inf (están en la carpeta /img) (?)

En el notebook de CV:

- Comentar gráfica.

- Comentar mejores parámetros y resultado en test.

## Kernel PCA

*Vary the parameters of the kernel and comment on the behavior of the projections onto the first two KPCA components for the different values considered (e.g.* $\gamma \in \{0.02, 0.2, 2.0, 20.0, 200.0, 2000.0\}$*. In particular,*

1. *What is the behavior in the limit in which the width of the kernel approaches* $\infty$*. Explain why one should expect such behavior.*
2. *What is the behavior in the limit in which the width of the kernel approaches* $0$*. Explain why one should expect such behavior.*

First of all, to study the limit behavior of the projections we are using the explicit RBF kernel formula:

$$\mathcal{K}(x, y) = A \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right),$$

where $A$ is the output variance and $\sigma$ is the kernel's width. In our case, $A = 1.0$ and the parameter $\gamma$ is inversely proportional to $\sigma$. Using this formula, and given two fixed input values $x \neq y$, we have:

$$\lim_{\gamma \to 0^+} = \lim_{\sigma \to \infty} \mathcal{K}(x, y) = A = 1.$$

$$\lim_{\gamma \to \infty} = \lim_{\sigma \to 0^+} \mathcal{K}(x, y) = 0.$$

Consider a generic set of points $\boldsymbol{X}$, with $N$ samples, and observe that in the code we only ever compute $\mathcal{K}(\boldsymbol{X}, \boldsymbol{X})$, that is, the test set is the same as the training set.

Let us begin by analyzing the first limit. In this case, if $x = y$ we have $\mathcal{K}(x, y) = A = 1$ as well, so all in all we have:

$$\lim_{\gamma \to 0^+} \mathcal{K}(\boldsymbol{X}, \boldsymbol{X}) = \mathbf{1},$$

where $\mathbf{1}$ is a matrix with the same dimension as $K \equiv \mathcal{K}(\boldsymbol{X}, \boldsymbol{X})$ containing all ones. Now observe that when we center this kernel, the expression in the limit is:

$$\hat{K} = \mathbf{1} - (2/N)\mathbf{1}\mathbf{1} + (1/N^2)\mathbf{1}\mathbf{1}\mathbf{1} = \mathbf{0},$$

that is, we get the null matrix. As a result, the limit case reduces to the matrix operation

$$\mathbf{0} \begin{pmatrix} e_1 & \dots e_N \end{pmatrix} = \mathbf{0},$$

where $e_1, \dots, e_N$ are the generated eigen-vectors of a null matrix, i.e, the usual basis. When $\hat{K}$ is not entirely a null matrix but it is near to it, we may see it as

$$\hat{K} = \epsilon \mathbf{1}, \quad \epsilon << 1.$$

The normalized eigen-vectors of such matrix are the same for each value of $\epsilon$, as a result, the matrix multiplication consists of scaling via $\epsilon$ the desired projection.

$$\hat{K} \begin{pmatrix} e_1 & \dots e_N \end{pmatrix} = \epsilon \mathbf{1} \begin{pmatrix} e_1 & \dots e_N \end{pmatrix} = \epsilon \begin{pmatrix} \|e_1\| & \dots \|e_N\| \end{pmatrix}.$$

This can be seen in the animation where as the kernel length tends to 0, the projection is preserved but reduced in scale.

Next we consider the second case, where the kernel tends to 0 on any pair of non equal points. This does not apply when computing $\lim_{\gamma \to \infty} \mathcal{K}(x, x)$, as $\mathcal{K}(x, x) = 1$ is constant for any given value of $\gamma$. Considering this, the kernel matrix converges to

$$\lim_{\gamma \to \infty} \mathcal{K}(\boldsymbol{X}, \boldsymbol{X}) = \boldsymbol{I},$$

where $\boldsymbol{I}$ is the identity matrix of order $N$. In this case, centering the kernel amounts to a small perturbation of $1/N$, so it can be ignored for our purposes. The situation in this case is this: the application of the kernel to any set of points results in an identity matrix, which has all of its eigenvalues equal to 1, and its eigenvector can be any orthonormal basis in $\mathbb{R}^N$. There is no need for such base to be the usual basis of $\mathbb{R}^N$, but it is the case when using Numpy's `eigh` function.

As a result, the kernel method performs the following matrix operation in the limit:

$$\mathcal{K}(\boldsymbol{X}, \boldsymbol{X}) \begin{pmatrix} e_1 & \dots & e_N \end{pmatrix} = \boldsymbol{I}\boldsymbol{I} = \boldsymbol{I}.$$

Given this, plotting a projection over the first two principal components leads to the degenerate case of projecting a point in $(1, 0)$ and another point in $(0, 1)$, while the rest are mapped to the origin. This phenomenon can be seen if we set a large enough value of $\gamma$.

However, there is an interesting behaviour of the `eigh` function when its argument is almost an identity matrix, but not exactly. The returned eigenvalues are roughly equal to 1 and the eigenvectors are nearly an orthonormal basis of $\mathbb{R}^N$, but they are not close to being the usual base. As a result, the plot leads to the projection over $\mathbb{R}^2$ of a basis in $\mathbb{R}^N$, which corresponds to the "spiked" appearance of the projection, in which the lines that are formed seem to be the projection of perpendicular axes from a higher dimensional space.