MAP569 Machine Learning II

PC7 : Kernel PCA

Exercise 1. Refresher on matrices

1. Let $\mathbf{A}$ be a $n \times d$ matrix with real entries. Show that $\mathrm{Im}(\mathbf{A}) = \mathrm{Im}(\mathbf{A}\mathbf{A}^T)$.
Solution.

First note that $\mathbf{A}\mathbf{A}^T x = 0$ implies $< \mathbf{A}^T x, \mathbf{A}^T x >= 0$ so that $\mathbf{A}^T x = 0$. The converse is obvious. Therefore, $\mathrm{Ker}(\mathbf{A}\mathbf{A}^T) = \mathrm{Ker}(\mathbf{A}^T)$. And using that $\mathrm{Ker}(B^T) = (\mathrm{Im}(B))^\perp$, we deduce that $\mathrm{Im}(\mathbf{A}\mathbf{A}^T)^\perp = \mathrm{Im}(\mathbf{A})^\perp$, which concludes the proof. □

2. Let $\{U_k\}_{1 \leq k \leq r}$ be a family of r orthonormal vectors of $\mathbb{R}^d$. Show that $\sum_{k=1}^r U_k U_k^T$ is the matrix associated with the orthogonal projection onto $\mathbf{H} = \{\sum_{k=1}^r \alpha_k U_k\ ;\ \alpha_1, \ldots, \alpha_r \in \mathbb{R}\}$. Deduce that if $\mathbf{A}$ is a $n \times d$ matrix with real entries such that each column of $\mathbf{A}$ is in H, then,

$$\left(\sum_{k=1}^r U_k U_k^T\right) \mathbf{A} = \mathbf{A}\,.$$

Solution.

Let $\pi_{\mathbf{H}}(X)$ be the orthogonal projection of X onto $\mathbf{H}$. Since $\{U_k\}_{1 \leq k \leq r}$ is an orthonormal basis of $\mathbf{H}$,

$$\pi_{\mathbf{H}}(X) = \sum_{k=1}^r < X, U_k > U_k = \left(\sum_{k=1}^r U_k U_k^T\right) X\,.$$

This implies that for each $X \in \mathbf{H}$, $X = \left(\sum_{k=1}^r U_k U_k^T\right) X$. Since all the column vectors of A are in $\mathbf{H}$, this yields $\left(\sum_{k=1}^r U_k U_k^T\right) \mathbf{A} = \mathbf{A}$. □

Exercise 2. Kernel Principal Component Analysis

Principal Component Analysis

Principal component analysis is a multivariate technique which aims at analyzing the statistical structure of high dimensional dependent observations by representing data using orthogonal variables called *principal components*. Reducing the dimensionality of the data is motivated by several practical reasons such as improving computational complexity. Let $(X_i)_{1 \leqslant i \leqslant n}$ be i.i.d. random variables in $\mathbb{R}^d$ and consider the matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$ such that the i-th row of $\mathbf{X}$ is the observation X_i^T. In this exercise, it is assumed that data are preprocessed so that the columns of $\mathbf{X}$ are centered. This means that for all $1 \leqslant k \leqslant d$, $\sum_{i=1}^n X_{i,k} = 0$. Let $\mathbf{\Sigma}_n$ be the empirical covariance matrix:

$$\mathbf{\Sigma}_n = n^{-1} \sum_{i=1}^n X_i X_i^T\,.$$

Principal Component Analysis aims at reducing the dimensionality of the observations $(X_i)_{1 \leqslant i \leqslant n}$ using a *compression* matrix $\mathbf{U} \in \mathbb{R}^{d \times p}$ with orthonormal columns with $p \leqslant d$ so that for each $1 \leqslant i \leqslant n$, $\mathbf{U}^T X_i$ ia a low dimensional representation of X_i. The original observation may then be partially recovered using $\mathbf{U} \in \mathbb{R}^{d \times p}$. Principal Component Analysis computes $\mathbf{U}$ using the least squares approach:

$$\mathbf{U}_\star \in \underset{U \in \mathbb{R}^{d \times p}}{\mathrm{argmin}} \sum_{i=1}^n \|X_i - \mathbf{U}\mathbf{U}^T X_i\|^2\,,$$

1. Prove that for all $\mathbb{R}^{n\times d}$ matrix $\mathbf{A}$ with rank r, there exist $\sigma_1 \geqslant \ldots \geqslant \sigma_r > 0$ such that

$$\mathbf{A} = \sum_{k=1}^{r} \sigma_k u_k v_k^T \,,$$

where $\{u_1, \ldots, u_r\} \subset \mathbb{R}^n$ and $\{v_1, \ldots, v_r\} \subset \mathbb{R}^d$ are two families of orthonormal vectors. The vectors $\{u_1, \ldots, u_r\}$ (resp. $\{v_1, \ldots, v_r\}$) are the left-singular (resp. right-singular) vectors associated with $\{\sigma_1, \ldots, \sigma_r\}$, the singular values of $\mathbf{A}$.
Solution.

Since the matrix $\mathbf{A}\mathbf{A}^T$ is positive semidefinite, its spectral decomposition is given by

$$\mathbf{A}\mathbf{A}^T = \sum_{k=1}^{r} \lambda_k u_k u_k^T \,,$$

where $\lambda_1 \geqslant \ldots \geqslant \lambda_r > 0$ are the nonzero eigenvalues of $\mathbf{A}\mathbf{A}^T$ and $\{u_1, \ldots, u_r\}$ is an orthonormal family of $\mathbb{R}^n$. For all $1 \leqslant k \leqslant r$, define $v_k = \lambda_k^{-1/2}\mathbf{A}^T u_k$ so that

$$\|v_k\|^2 = \lambda_k^{-1}\langle \mathbf{A}^T u_k; \mathbf{A}^T u_k\rangle = \lambda_k^{-1} u_k^T \mathbf{A}\mathbf{A}^T u_k = 1 \,,$$
$$\mathbf{A}^T\mathbf{A} v_k = \lambda_k^{-1/2}\mathbf{A}^T\mathbf{A}\mathbf{A}^T u_k = \lambda_k v_k \,.$$

On the other hand, for all $1 \leqslant k \neq j \leqslant r$, $\langle v_k; v_j\rangle = \lambda_k^{-1/2}\lambda_j^{-1/2} u_k^T \mathbf{A}\mathbf{A}^T u_j = \lambda_k^{-1/2}\lambda_j^{1/2} u_k' u_j = 0$. Therefore, $\{v_1, \ldots, v_r\}$ is an orthonormal family of eigenvectors of $\mathbf{A}^T\mathbf{A}$ associated with the eigenvalues $\lambda_1 \geqslant \ldots \geqslant \lambda_r > 0$. Define, for all $1 \leqslant k \leqslant r$, $\sigma_k = \lambda_k^{1/2}$ which yields

$$\sum_{k=1}^{r} \sigma_k u_k v_k^T = \sum_{k=1}^{r} u_k u_k^T \mathbf{A} = \left(\sum_{k=1}^{r} u_k u_k^T\right)\mathbf{A} \,.$$

As $\{u_1, \ldots, u_r\}$ is an orthonormal family, $\mathbf{U}\mathbf{U}^T = \sum_{k=1}^{r} u_k u_k^T$ is the orthogonal projection onto the range$(\mathbf{A}\mathbf{A}^T) = \mathrm{range}(\mathbf{A})$ which implies

$$\sum_{k=1}^{r} \sigma_k u_k v_k^T = \left(\sum_{k=1}^{r} u_k u_k^T\right)\mathbf{A} = \mathbf{A} \,.$$

□

If $\mathbf{U}$ denotes the $\mathbb{R}^{n\times r}$ matrix with columns given by $\{u_1, \ldots, u_r\}$ and $\mathbf{V}$ denotes the $\mathbb{R}^{d\times r}$ matrix with columns given by $\{v_1, \ldots, v_r\}$, then the singular value decomposition of $\mathbf{A}$ may also be written as

$$\mathbf{A} = \mathbf{U}\mathbf{D}_r\mathbf{V}^T \,,$$

where $\mathbf{D}_r = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$. Then, $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$ are positive semidefinite such that

$$\mathbf{A}^T\mathbf{A} = \mathbf{V}\mathbf{D}_r^2\mathbf{V}^T \quad \text{and} \quad \mathbf{A}\mathbf{A}^T = \mathbf{U}\mathbf{D}_r^2\mathbf{U}^T \,.$$

In the framwework of this exercise, $n\mathbf{\Sigma}_n = \mathbf{X}^T\mathbf{X}$ so that diagonalizing $n\mathbf{\Sigma}_n$ is equivalent to computing the singular value decomposition of $\mathbf{X}$.

2. Prove that solving the PCA least squares optimization problem boils down to computing

$$\mathbf{U}_\star \in \underset{\mathbf{U}\in\mathbb{R}^{d\times p}\,,\,\mathbf{U}^T\mathbf{U}=\mathbf{I}_p}{\mathrm{argmax}} \{\mathrm{trace}(\mathbf{U}^T\mathbf{\Sigma}_n\mathbf{U})\} \,.$$

Solution.

Let $\mathbf{U} \in \mathbb{R}^{d\times p}$ be such that $\mathbf{U}^T\mathbf{U} = \mathbf{I}_p$. Then,

$$\begin{aligned}\sum_{i=1}^n \|X_i - \mathbf{U}\mathbf{U}^T X_i\|^2 &= \sum_{i=1}^n \|X_i\|^2 + \sum_{i=1}^n \|\mathbf{U}\mathbf{U}^T X_i\|^2 - 2\sum_{i=1}^n \langle X_i ; \mathbf{U}\mathbf{U}^T X_i\rangle\,,\\ &= \sum_{i=1}^n \|X_i\|^2 + \sum_{i=1}^n X_i^T\mathbf{U}\mathbf{U}^T X_i - 2\sum_{i=1}^n X_i^T\mathbf{U}\mathbf{U}^T X_i\,,\\ &= \sum_{i=1}^n \|X_i\|^2 - \sum_{i=1}^n X_i^T\mathbf{U}\mathbf{U}^T X_i\,,\\ &= \sum_{i=1}^n \|X_i\|^2 - \mathrm{trace}(\mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U})\,.\end{aligned}$$

□

3. Let $\{\vartheta_1, \ldots, \vartheta_d\}$ be orthonormal eigenvectors associated with the eigenvalues $\lambda_1 \geqslant \ldots \geqslant \lambda_d$ of $\mathbf{\Sigma}_n$. Prove that a solution to this problem is given by the matrix $\mathbf{U}_\star$ with columns $\{\vartheta_1, \ldots, \vartheta_p\}$.
Solution.

Let $\mathbf{\Sigma}_n = \mathbf{V}\mathbf{D}_n\mathbf{V}^T$ be the spectral decomposition of $\mathbf{\Sigma}_n$ where $\mathbf{D}_n = \mathrm{Diag}(\lambda_1, \ldots, \lambda_d)$ and $\mathbf{V} \in \mathbb{R}^{d\times d}$ is a matrix with orthonormal columns $\{\vartheta_1, \ldots, \vartheta_d\}$. For all $\mathbf{U} \in \mathbb{R}^{d\times p}$ matrix with orthonormal columns define $\mathbf{B} = \mathbf{V}^T\mathbf{U}$ so that, as $\mathbf{V} \in \mathbb{R}^{d\times d}$ is an orthogonal matrix,

$$\mathbf{V}\mathbf{B} = \mathbf{V}\mathbf{V}^T\mathbf{U} = \mathbf{U} \quad \text{and} \quad \mathbf{U}^T\mathbf{\Sigma}_n\mathbf{U} = \mathbf{B}^T\mathbf{V}^T\mathbf{V}\mathbf{D}_n\mathbf{V}^T\mathbf{V}\mathbf{B} = \mathbf{B}^T\mathbf{D}_n\mathbf{B}\,.$$

Therefore,

$$\mathrm{Trace}(\mathbf{U}^T\mathbf{\Sigma}_n\mathbf{U}) = \mathrm{Trace}(\mathbf{B}^T\mathbf{D}_n\mathbf{B}) = \sum_{i=1}^d \lambda_i \sum_{j=1}^p b_{i,j}^2\,. \tag{1}$$

On the other hand,

$$\mathbf{B}^T\mathbf{B} = \mathbf{U}^T\mathbf{V}\mathbf{V}^T\mathbf{U} = \mathbf{U}^T\mathbf{U} = I_p\,,$$

so that the columns of $\mathbf{B}$ are orthonormal and

$$\sum_{i=1}^d\sum_{j=1}^p b_{i,j}^2 = p\,.$$

Hence, introducing for all $1 \leqslant i \leqslant d$, $\alpha_i = \sum_{j=1}^p b_{i,j}^2$, by (1),

$$\mathrm{Trace}(\mathbf{U}^T\mathbf{\Sigma}_n\mathbf{U}) = \sum_{i=1}^d \alpha_i\lambda_i\,,$$

with, for all $1 \leqslant i \leqslant d$, $\alpha_i \in [0,1]$ and $\sum_{i=1}^d \alpha_i = p$. As $\lambda_1 \geqslant \lambda_2 \geqslant \ldots, \lambda_d$,

$$\mathrm{Trace}(\mathbf{U}^T\mathbf{\Sigma}_n\mathbf{U}) \leqslant \sum_{i=1}^p \lambda_i\,.$$

Indeed, the function $f_d : (\alpha_1, \ldots, \alpha_d) \mapsto \sum_{i=1}^d \alpha_i\lambda_i$ is maximized under the constraints $\alpha_i \in [0,1]$ and $\sum_{i=1}^d \alpha_i = p$ by $(\alpha_i^*)_{1\leqslant i\leqslant d}$ such that $\alpha_1^* = \ldots = \alpha_p^* = 1$. Assume that $(\alpha_1, \ldots, \alpha_d)$ is such that there exists $1 \leqslant j_0 \leqslant p$ such that $\alpha_{j_0} < 1$. Then, $\sum_{j=p+1}^d \alpha_j \geqslant 1 - \alpha_{j_0}$ and we may write, as $\lambda_{j_0} \geqslant \lambda_{p+1} \geqslant \ldots \geqslant \lambda_d$,

$$f_d : (\alpha_1, \ldots, \alpha_d) \leqslant \sum_{i=1, i\neq j_0}^p \alpha_i\lambda_i + \lambda_{j_0} + \sum_{i=p+1}^d \tilde{\alpha}_i\lambda_i\,,$$

where $(\tilde{\alpha}_i)_{p+1\leqslant i\leqslant d}$ are in $[0,1]$ and such that $\sum_{i=1,i\neq j_0}^p \alpha_i + 1 + \sum_{i=p+1}^d \tilde{\alpha}_i = p$.

As the columns of $\mathbf{U}_\star$ are $\{\vartheta_1, \ldots, \vartheta_p\}$, for all $1 \leqslant i \leqslant d$ and $1 \leqslant j \leqslant p$, $b_{i,j} = \langle \vartheta_i ; \vartheta_j\rangle = \delta_{i,j}$. Therefore, for all $1 \leqslant i \leqslant d$, $\sum_{j=1}^p b_{i,j}^2 = 1$ and

$$\mathrm{Trace}(\mathbf{U}_\star^T\mathbf{\Sigma}_n\mathbf{U}_\star) = \sum_{i=1}^p \lambda_i\,,$$

which completes the proof. □

4. For any dimension $1 \leqslant p \leqslant d$, let $\mathcal{F}_d^p$ be the set of all vector subpaces of $\mathbb{R}^d$ with dimension p. Consider the linear span V_d defined as

$$V_p \in \underset{V \in \mathcal{F}_d^p}{\operatorname{argmin}} \sum_{i=1}^{n} \|X_i - \pi_V(X_i)\|^2 \,,$$

where π_V is the orthogonal projection onto the linear span V. Prove that $V_1 = \operatorname{span}\{v_1\}$ where

$$v_1 \in \underset{v \in \mathbb{R}^d \,;\, \|v\|=1}{\operatorname{argmax}} \sum_{i=1}^{n} \langle X_i, v\rangle^2 \,.$$

Solution.

Write $V_1 = \operatorname{span}\{v_1\}$ for $v_1 \in \mathbb{R}^d$ such that $\|v_1\| = 1$. Then,

$$\begin{aligned}
\sum_{i=1}^{n} \|X_i - \pi_{V_1}(X_i)\|^2 &= \sum_{i=1}^{n} \|X_i - \langle X_i; v_1\rangle v_1\|^2 \,,\\
&= \sum_{i=1}^{n} \left(\|X_i\|^2 - 2\langle X_i; \langle X_i; v_1\rangle v_1\rangle + \|\langle X_i; v_1\rangle v_1\|^2\right) \,,\\
&= \sum_{i=1}^{n} \left(\|X_i\|^2 - \langle X_i; v_1\rangle^2\right) .
\end{aligned}$$

Consequently, V_1 is a solution if and only if v_1 is solution to:

$$v_1 \in \underset{v \in \mathbb{R}^d \,;\, \|v\|=1}{\operatorname{argmax}} \sum_{i=1}^{n} \langle X_i, v\rangle^2 \,.$$

□

5. For all $2 \leqslant p \leqslant d$, following the same steps, prove that a solution to the optimization problem is given by $V_p = \operatorname{span}\{v_1, \ldots, v_p\}$ where

$$v_1 \in \underset{v \in \mathbb{R}^d \,;\, \|v\|=1}{\operatorname{argmax}} \sum_{i=1}^{n} \langle X_i, v\rangle^2 \quad \text{and for all } \; 2 \leqslant k \leqslant p \,, \;\; v_k \in \underset{\substack{v \in \mathbb{R}^d \,;\, \|v\|=1 \,;\\ v \perp v_1, \ldots, v \perp v_{k-1}}}{\operatorname{argmax}} \sum_{i=1}^{n} \langle X_i, v\rangle^2 \,. \quad (2)$$

Solution.

Write $V_p = \operatorname{span}\{v_1, \ldots, v_p\}$ where $\{v_1, \ldots, v_p\}$ is an orthonormal family. Then,

$$\sum_{i=1}^{n} \|X_i - \pi_{V_p}(X_i)\|^2 = \sum_{i=1}^{n} \|X_i - \sum_{k=1}^{p} \langle X_i; v_k\rangle v_k\|^2 = \sum_{i=1}^{n} \left(\|X_i\|^2 - \sum_{k=1}^{p} \langle X_i; v_k\rangle^2\right) .$$

$(v_1, \ldots, v_p)$ is therefore solution to

$$v = (v_1, \ldots, v_p) \in \operatorname{argmax} \sum_{k=1}^{p} \sum_{i=1}^{n} \langle X_i; v_k\rangle^2 \,.$$

The additive form of the function to be maximized allows to build the orthonormal basis of V_p sequentially as claimed. □

6. Prove that the vectors $\{v_1, \ldots, v_k\}$ defined by (2) can be chosen as the orthonormal eigenvectors associated with the k largest eigenvalues of the empirical covariance matrix $\mathbf{\Sigma}_n$.
Solution.

Note that for all $v \in \mathbb{R}^d$ such that $\|v\| = 1$,

$$\frac{1}{n} \sum_{i=1}^{n} \langle X_i, v\rangle^2 = \frac{1}{n} \sum_{i=1}^{n} (v^T X_i)(X_i^T v) = v^T \mathbf{\Sigma}_n v \,.$$

As $(\vartheta_i)_{1\leqslant i\leqslant d}$ are the orthonormal eigenvectors associated with the eigenvalues $\lambda_1 \geqslant \ldots \geqslant \lambda_d \geqslant 0$ of Σ_n. Then,

$$\frac{1}{n}\sum_{i=1}^{n}\langle X_i, v\rangle^2 = v^T\left(\sum_{i=1}^{d}\lambda_i\vartheta_i\vartheta_i^T\right)v = \sum_{i=1}^{d}\lambda_i\langle v,\vartheta_i\rangle^2 \leqslant \lambda_1\sum_{i=1}^{d}\langle v,\vartheta_i\rangle^2$$

and, as $(\vartheta_i)_{1\leqslant i\leqslant d}$ is an orthonormal basis of $\mathbb{R}^d$, $\sum_{i=1}^{d}\langle v,\vartheta_i\rangle^2 = \|v\|^2 = 1$. Therefore,

$$\frac{1}{n}\sum_{i=1}^{n}\langle X_i, v\rangle^2 \leqslant \lambda_1\,.$$

On the other hand, for all $2 \leqslant i \leqslant d$, $\langle\vartheta_1,\vartheta_i\rangle = 0$ and $\langle\vartheta_1,\vartheta_1\rangle = 1$ so that $\sum_{i=1}^{d}\lambda_i\langle\vartheta_1,\vartheta_i\rangle^2 = \lambda_1$ which proves that ϑ_1 is solution to (2).

Assume now that $v \in \mathbb{R}^d$ is such that $\|v\| = 1$ and for all $1 \leqslant j \leqslant k-1$, $\langle v;\vartheta_j\rangle = 0$ and write

$$\frac{1}{n}\sum_{i=1}^{n}\langle X_i, v\rangle^2 = \sum_{i=1}^{d}\lambda_i\langle v,\vartheta_i\rangle^2 \leq \lambda_k\sum_{i=k}^{d}\langle v,\vartheta_i\rangle^2 \leq \lambda_k\,,$$

since, as $(\vartheta_i)_{1\leqslant i\leqslant d}$ is an orthonormal basis of $\mathbb{R}^d$, $\sum_{i=1}^{d}\langle v,\vartheta_i\rangle^2 = \sum_{i=k}^{d}\langle v,\vartheta_i\rangle^2 = \|v\|^2 = 1$. On the other hand, for all $1 \leqslant i \leqslant d$, $i \neq k$, $\langle\vartheta_k,\vartheta_i\rangle = 0$ and $\langle\vartheta_k,\vartheta_k\rangle = 1$ so that $\sum_{i=1}^{d}\lambda_i\langle\vartheta_k,\vartheta_i\rangle^2 = \lambda_k$ which proves that ϑ_k is solution to (2).

Therefore, $V_p = \mathrm{span}\{\vartheta_1,\ldots\vartheta_p\}$ is a solution to (2) and, as $(\vartheta_i)_{1\leqslant i\leqslant p}$ is an orthonormal family, the projection matrix onto V_p is given by $\mathbf{U}_\star\mathbf{U}_\star^T$ where $\mathbf{U}_\star$ is a $\mathbb{R}^{d\times p}$ matrix with columns $\{\vartheta_1,\ldots\vartheta_p\}$. □

7. The orthonormal eigenvectors associated with the eigenvalues of Σ_n allow to define the principal components as follows. Then, as $V_d = \mathrm{span}\{\vartheta_1,\ldots,\vartheta_d\}$, for all $1 \leqslant i \leqslant n$,

$$\pi_{V_d}(X_i) = \sum_{k=1}^{d}\langle X_i,\vartheta_k\rangle\vartheta_k = \sum_{k=1}^{d}(X_i^T\vartheta_k)\vartheta_k = \sum_{k=1}^{d}c_k(i)\vartheta_k\,,$$

where for all $1 \leqslant k \leqslant d$, the k-th principal component is defined as $c_k = \mathbf{X}\vartheta_k$. Prove that $(c_1,\ldots,c_d)$ are orthogonal vectors.

Solution.

The k-th principal component is the vector whose components are the coordinates of each X_i, $1 \leqslant i \leqslant n$, relative to the basis $\{\vartheta_1,\ldots,\vartheta_d\}$ of V_d. For all $1 \leqslant i \neq j \leqslant d$,

$$\langle c_i, c_j\rangle = \vartheta_i^T\mathbf{X}^T\mathbf{X}\vartheta_j = \vartheta_i^T(n\Sigma_n)\vartheta_j = n\lambda_j\vartheta_i^T\vartheta_j = 0\,,$$

as $\{\vartheta_1,\ldots,\vartheta_d\}$ is an orthonormal family. □

Application to RKHS

Let $(X_i)_{1\leq i\leq n}$ be n observations in a general space $\mathcal{X}$ and $k : \mathcal{X}\times\mathcal{X}\to\mathbb{R}$ a positive kernel. $\mathcal{W}$ denotes the Reproducing Kernel Hilbert Space associated with k and for all $x \in \mathcal{X}$, $\phi(x)$ denotes the function $\phi(x) : y \to k(x,y)$. The aim is now to perform a PCA on $(\phi(X_1),\ldots,\phi(X_n))$. It is assumed that

$$\sum_{i=1}^{n}\phi(X_i) = 0\,.$$

Define

$$\mathbf{K} = (k(X_i,X_j))_{1\leqslant i,j\leqslant n}\,.$$

1. Prove that

$$f_1 = \operatorname*{argmax}_{f\in\mathcal{W}\,;\,\|f\|_{\mathcal{W}}=1}\sum_{i=1}^{n}\langle\phi(X_i),f\rangle_{\mathcal{W}}^2$$

may be written

$$f_1 = \sum_{i=1}^{n}\alpha_1(i)\phi(X_i)\,,\quad\text{where}\quad \alpha_1 = \operatorname*{argmax}_{\alpha\in\mathbb{R}^n\,;\,\alpha^T\mathbf{K}\alpha=1}\alpha^T\mathbf{K}^2\alpha\,.$$

Solution.

Any solution to the optimization problem lies in the vectorial subspace $V = \mathrm{span}\{\phi(X_i), \dots, \phi(X_n)\}$. Let $f = \sum_{i=1}^n \alpha(i)\phi(X_i)$ be such that $\|f\|_{\mathcal{W}} = 1$. Then,

$$\|f\|_{\mathcal{W}}^2 = \sum_{i,j=1}^n \alpha_i \alpha_j \langle \phi(X_i), \phi(X_j)\rangle_{\mathcal{W}} = \alpha^T \mathbf{K} \alpha \,.$$

On the other hand, $\langle \phi(X_i), f\rangle_{\mathcal{W}} = f(X_i) = [\mathbf{K}\alpha](i)$ so that,

$$\sum_{i=1}^n \langle \phi(X_i), f\rangle_{\mathcal{W}}^2 = \sum_{i=1}^n f^2(X_i) = \sum_{i=1}^n ([\mathbf{K}\alpha](i))^2 = (\mathbf{K}\alpha_1)^T \mathbf{K}\alpha_1 = \alpha^T \mathbf{K}^2 \alpha \,.$$

□

2. Prove that $\alpha_1 = \lambda_1^{-1/2} b_1$ where b_1 is the unit eigenvector associated with the largest eigenvalue λ_1 of $\mathbf{K}$.
Solution.

Let $\lambda_1 \geqslant \dots \geqslant \lambda_n \geq 0$ be the eigenvalues of $\mathbf{K}$ associated with the orthonormal basis of eigenvectors $(b_1, \dots, b_n)$. For any $\alpha \in \mathbb{R}^n$ such that $\alpha^T \mathbf{K} \alpha = 1$,

$$\alpha^T \mathbf{K}^2 \alpha = \alpha^T \left(\sum_{i=1}^n \lambda_i b_i b_i^T \right)^2 \alpha = \sum_{i=1}^n \lambda_i^2 \langle \alpha, b_i \rangle^2 \leqslant \lambda_1 \underbrace{\sum_{i=1}^n \lambda_i \langle \alpha, b_i \rangle^2}_{=1} = \lambda_1 \,,$$

as $\alpha^T \mathbf{K} \alpha = \sum_{i=1}^n \lambda_i \langle \alpha, b_i \rangle^2 = 1$. On the other hand,

$$\left(\lambda_1^{-1/2} b_1\right)^T \mathbf{K}^2 \left(\lambda_1^{-1/2} b_1\right) = \lambda_1^{-1} \sum_{i=1}^n \lambda_i^2 \langle b_1, b_i \rangle^2 = \lambda_1 \,.$$

Following the same steps, f_j may be written $f_j = \sum_{i=1}^n \alpha_j(i)\phi(x_i)$ with $\alpha_j = \lambda_j^{-1/2} b_j$. □

3. Write $H_d = \mathrm{span}\{f_1, \dots, f_d\}$. Prove that, for all $1 \leqslant i \leqslant n$,

$$\pi_{H_d}(\phi(X_i)) = \sum_{j=1}^d \lambda_j \alpha_j(i) f_j \,.$$

Solution.

Note first that the $(f_1, \dots, f_d)$ is an orthonormal family. Therefore,

$$\pi_{H_d}(\phi(X_i)) = \sum_{j=1}^d \langle \phi(X_i), f_j \rangle_{\mathcal{W}} f_j = \sum_{j=1}^d \langle \phi(X_i), \sum_{\ell=1}^n \alpha_j(\ell)\phi(X_\ell) \rangle_{\mathcal{W}} f_j = \sum_{j=1}^d [\mathbf{K}\alpha_j](i) f_j \,.$$

Therefore,

$$\pi_{H_d}(\phi(x_i)) = \sum_{j=1}^d \lambda_j^{-1/2} [\mathbf{K} b_j](i) f_j = \sum_{j=1}^d \lambda_j^{1/2} b_j(i) f_j = \sum_{j=1}^d \lambda_j \alpha_j(i) f_j \,.$$

□