1 Warm-up

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

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

2 Kernel Principal Component Analysis

2.1 Principal Component Analysis

Let $(X_i)_{1\leqslant i\leqslant n}$ be i.i.d. random variables in $\mathbb{R}^d$ and consider the matrix $X \in \mathbb{R}^{n\times d}$ such that the i-th row of X is the observation $X_i^\top$. In this exercise, it is assumed that data are preprocessed so that the columns of X are centered. This means that for all $1 \leqslant k \leqslant d$, $\sum_{i=1}^n X_{i,k} = 0$. Let Σ_n be the empirical covariance matrix:
$$\Sigma_n = n^{-1}\sum_{i=1}^n X_i X_i^\top\,.$$
Principal Component Analysis aims at reducing the dimensionality of the observations $(X_i)_{1\leqslant i\leqslant n}$ using a *compression* matrix $U \in \mathbb{R}^{d\times p}$ with orthonormal columns with $p \leqslant d$ so that for each $1 \leqslant i \leqslant n$, $U^\top X_i$ ia a low dimensional representation of X_i. The original observation may then be partially recovered using $U \in \mathbb{R}^{d\times p}$. Principal Component Analysis computes U using the least squares approach:
$$U_\star \in \operatorname*{argmin}_{U\in\mathbb{R}^{d\times p}} \sum_{i=1}^n \|X_i - UU^\top X_i\|_2^2\,,$$

1. Prove that for all $\mathbb{R}^{n\times d}$ matrix A with rank r, there exist $\sigma_1 \geqslant \ldots \geqslant \sigma_r > 0$ such that
$$A = \sum_{k=1}^r \sigma_k u_k v_k^\top\,,$$
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 A.

2. Prove that solving the PCA least squares optimization problem boils down to computing
$$U_\star \in \operatorname*{argmax}_{U\in\mathbb{R}^{d\times p}\,,\,U^\top U=\mathrm{I}_p} \{\mathrm{trace}(U^\top \Sigma_n U)\}\,.$$

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

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
$$V_p \in \underset{V \in \mathcal{F}_d^p}{\operatorname{argmin}} \sum_{i=1}^n \|X_i - \pi_V(X_i)\|_2^2 \,,$$
where π_V is the orthogonal projection onto the linear span V. Prove that $V_1 = \operatorname{span}\{v_1\}$ where
$$v_1 \in \underset{v \in \mathbb{R}^d \,;\, \|v\|_2 = 1}{\operatorname{argmax}} \sum_{i=1}^n \langle X_i, v\rangle^2 \,.$$
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
$$v_1 \in \underset{v \in \mathbb{R}^d \,;\, \|v\| = 1}{\operatorname{argmax}} \sum_{i=1}^n \langle X_i, v\rangle^2 \quad \text{and for all } 2 \leqslant k \leqslant p \,, \;\; v_k \in \underset{\substack{v \in \mathbb{R}^d \,;\, \|v\| = 1\,; \\ v \perp v_1, \ldots, v \perp v_{k-1}}}{\operatorname{argmax}} \sum_{i=1}^n \langle X_i, v\rangle^2 \,. \quad (1)$$
6. Prove that the vectors $\{v_1, \ldots, v_k\}$ defined by (1) can be chosen as the orthonormal eigenvectors associated with the k largest eigenvalues of the empirical covariance matrix Σ_n.
7. The orthonormal eigenvectors associated with the eigenvalues of Σ_n allow to define the principal components as follows. Then, as $V_d = \operatorname{span}\{\vartheta_1, \ldots, \vartheta_d\}$, for all $1 \leqslant i \leqslant n$,
$$\pi_{V_d}(X_i) = \sum_{k=1}^d \langle X_i, \vartheta_k\rangle \vartheta_k = \sum_{k=1}^d (X_i^\top \vartheta_k)\vartheta_k = \sum_{k=1}^d c_k(i)\vartheta_k \,,$$
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.

2.2 Application to RKHS

Let $(X_i)_{1 \le i \le n}$ be n observations in a general space $\mathcal{X}$ and $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ a positive function. We assume that k is symmetric and that for all $n \geqslant 1$, $(a_i)_{1 \leqslant i \leqslant n} \in \mathbb{R}^n$ and $(x_i)_{1 \leqslant i \leqslant n} \in \mathcal{X}^n$, $\sum_{1 \leqslant i,j \leqslant n} a_i a_j k(x_i, x_j) \geqslant 0$. For all $x \in \mathcal{X}$, $\phi(x)$ denotes the function $\phi(x) : y \to k(x, y)$.

Let $\mathcal{W}$ be a Hilbert space of real-valued functions defined on $\mathcal{X}$, endowed with an inner product denoted by $\langle \cdot, \cdot \rangle_{\mathcal{W}}$, and such that for all $x \in \mathcal{X}$, $\phi(x) \in \mathcal{W}$ and for all $f \in \mathcal{W}$ and all $x \in \mathcal{X}$, $f(x) = \langle f, \phi(x)\rangle_{\mathcal{W}}$. 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
$$K = (k(X_i, X_j))_{1 \leqslant i,j \leqslant n} \,.$$

1. Prove that
$$f_1 = \underset{f \in \mathcal{W} \,;\, \|f\|_{\mathcal{W}} = 1}{\operatorname{argmax}} \sum_{i=1}^n \langle \phi(X_i), f\rangle_{\mathcal{W}}^2$$
may be written
$$f_1 = \sum_{i=1}^n \alpha_1(i)\phi(X_i) \,, \quad \text{where} \quad \alpha_1 = \underset{\alpha \in \mathbb{R}^n \,;\, \alpha^T K \alpha = 1}{\operatorname{argmax}} \alpha^\top K^2 \alpha \,.$$
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 K.
3. Write $H_d = \operatorname{span}\{f_1, \ldots, f_d\}$. Prove that, for all $1 \leqslant i \leqslant n$,
$$\pi_{H_d}(\phi(X_i)) = \sum_{j=1}^d \lambda_j \alpha_j(i) f_j \,.$$