3.4.4. Algorithm: (Nonlinear / Kernel PCA)

Input: sample pts $x^{(1)}, \dots, x^{(n)} \in \mathbb{R}^p$ and
a feature map $\phi : \mathbb{R}^p \to \mathbb{R}^D$ $(D \gg p)$
or kernel function $k(x, y)$ $(= \phi^T(x)\phi(y))$

Step 1: Compute the non-centered kernel matrix

$$\Phi^T\Phi = k(x^{(i)}, x^{(j)}) = \phi^T(x^{(i)})\phi(x^{(j)})$$

and centered kernel matrix $H\Phi^T\Phi H$.

Step 2: Compute the EVD of $H\Phi^T\Phi H$:

$$H\Phi^T\Phi H = \underbrace{\begin{pmatrix} | & & | \\ u^{(1)} & \cdots & u^{(n)} \\ | & & | \end{pmatrix}}_{\text{orthogonal } U} \underbrace{\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}}_{\text{diagonal}} \underbrace{\begin{pmatrix} — u^{(1)} — \\ \vdots \\ — u^{(n)} — \end{pmatrix}}_{U^T}$$

Step 3: Compute $v^{(i)} = \frac{1}{\sqrt{\lambda_i}} \Phi u^{(i)}$, $i = 1, \dots, d$

Output: principal directions $v^{(1)}, \dots, v^{(d)}$

Remark: For Step 3, if only $k(x,y)$ is available, we can directly compute the principal component of a sample pt $\phi(x)$ as

$$\left(v^{(i)T} \phi(x)\right) v^{(i)} = \left(\frac{1}{\sqrt{\lambda_i}} u^{(i)T} \underbrace{\Phi^T \phi(x)}_{= k(x^{(i)}, x)}\right) v^{(i)}$$

Summary of this section:

- What is nonlinear / Kernel PCA ? : Def
- Why nonlinear / Kernel PCA ? : Motivation
- How to implement ? : Algorithm.

Reference: "Generalized PCA" (in the syllabus) Section 4.1.