Kernel PCA

$\boxed{\text{Consider standard PCA:}}$

$X \in \mathbb{R}^{N \times D}$ $\qquad \frac{1}{N} X^T X u = \lambda u$ (assumption: X is centered already

$$\Rightarrow \quad u = \frac{1}{\lambda N} X^T X u = X^T \left(\frac{1}{\lambda N} X u \right) = X^T \alpha$$

What is α?

$$\Rightarrow \quad \frac{1}{N} X^T X \, X^T \alpha = \lambda \, X^T \alpha$$

$$\Rightarrow \quad \frac{1}{N} X \cdot X^T X X^T \alpha = \lambda \, X X^T \alpha$$

$$\Rightarrow \quad \frac{1}{N} (X \cdot X^T)(X \cdot X^T) \alpha = \lambda (X X^T) \alpha$$

Replace $X \cdot X^T$ (linear kernel matrix) with any kernel matrix

$$\Rightarrow \quad \frac{1}{N} K \cdot K \alpha = \lambda K \alpha$$

$$\Rightarrow \quad \frac{1}{N} K \alpha = \lambda \alpha \quad \Rightarrow \quad K \alpha = \lambda N \alpha$$

$\boxed{\text{Ie, } \alpha \text{ is the eigenvector of kernel matrix !!!}}$

What is u then?

$$u = X^T \alpha = X^T \cdot \alpha$$

use kernel matrix

this is still the original feature.

However the projection coordinates

$$z = x^T u = x \cdot X^T \alpha \quad \text{is}$$

$$= \underbrace{[\, k(x_1, x), \; k(x_2, x) \cdots \; k(x_N, x) \,]}_{\text{computed using kernel function}} \alpha$$

so we don't have to figure out what u is to compute z !!!

Subtle issue 1

In PCA, we have assumed X is centered.

In kernel-PCA, how are we going to assure that?

use

$$K \leftarrow K - 1_N \cdot K - K 1_N + 1_N K 1_N$$

$$1_N = \frac{1}{N}\begin{bmatrix} 1 & 1 & \cdots & \cdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$ ← all $N \times N$ "1"

how this is derived?

To centralize X, we use

$$\cancel{(I - \tfrac{1}{N}} \quad (I - 1_N)X$$

It's kernel matrix is

$$(I - 1_N) X X^T (I - 1_N)$$

$$\Rightarrow (I - 1_N) K (I - 1_N)$$

$$\xRightarrow{\text{expand}} K - 1_N \cdot K - K 1_N + 1_N K 1_N$$

Subtle issue 2

To ensure $\|u\|_2^2 = 1$, we need to rescale α by $\frac{1}{\sqrt{\lambda N}}$

(see eq. (12.81) in P.R.M.L)