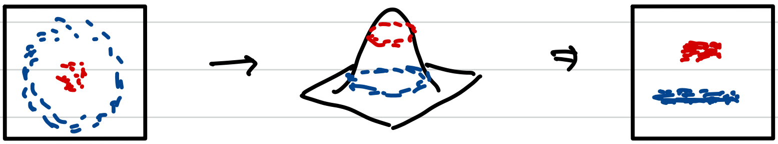
Kernel PCA

: Original space 에서의 PCA의 장점을 극대화시키기 위해 feature space로 확장 *

ex) .



* Assumption. : the projected new features have zero mean, i.e.

$$\frac{1}{N}\sum_{i=1}^{n}\phi(x_i)=0 \quad \left(\overset{\text{let}}{=\!=} m^{\phi}\right).$$

* The Covariance matrix of the projected features is $M\times M$. Calculated by

$$C^{\phi}=\frac{1}{N}\sum_{i=1}^{n}\left(\phi(x_i)-\overset{=0}{m^{\phi}}\right)\left(\phi(x_i)-\overset{=0}{m^{\phi}}\right)=\frac{1}{N}\sum_{i=1}^{n}\phi(x_i)\phi(x_i)^T$$

Its eigenvalues and eigenvectors are given by

$$C^{\phi}v_k=\lambda_k v_k$$

* From equations above, $\underbrace{\frac{1}{N}\sum_{i=1}^{n}\phi(x_i)(\phi(x_i)^T}_{C} v_k)=\lambda_k v_k.$

which can be written as

$$v_k=\sum_{i=1}^{n}\alpha_{ki}\phi(x_i)$$

By substituting v_k,

$$\frac{1}{N}\sum_{i=1}^{n}\phi(x_i)\phi(x_i)^T\sum_{j=1}^{N}\alpha_{kj}\phi(x_j)=\lambda_k\sum_{i=1}^{N}\alpha_{ki}\phi(x_i) \quad \cdots (*)$$

* If we define the kernel function.

$$K(x_i,x_j)=\phi(x_i)^T\phi(x_j)$$

* And multiply both side of $(*)$ by $\phi(x_l)^T$

$$\frac{1}{N}\sum_{i=1}^{n}\phi(x_l)^T\phi(x_i)\sum_{j=1}^{n}\alpha_{kj}\phi(x_i)^T\phi(x_j)=\lambda_k\sum_{j=1}^{N}\alpha_{kj}\phi(x_l)^T\phi(x_j)$$

$$\Rightarrow \frac{1}{N}\sum_{i=1}^{N}K(x_l,x_i)\sum_{j=1}^{N}\alpha_{kj}K(x_i,x_j)=\lambda_k\sum_{j=1}^{N}\alpha_{kj}K(x_l,x_j)$$

Matrix notation $\quad K^{\cancel{2}}\alpha_k=\lambda_k N\cancel{K}\alpha_k \;\Rightarrow\; K\alpha_k=\lambda_k N\alpha_k$

* The eigenvector problem becomes

$$K\alpha_k = \lambda_k N \alpha_k$$

eigenvector of $K_{n \times n}$ *

* And the resulting kernel PC can be calculated using

$$y_k(x) = \phi(x)^T V_k = \sum_{i=1}^{N} \alpha_{ki} K(x, x_i)$$ ϕ의 형태 고려할 필요X

* If the projected dataset does not have zero mean. Gram matrix.

$$\tilde{K} = (I - 1_N) K (I - 1_N)$$

$$= K - 1_N K - K 1_N + 1_N K 1_N$$

Standard PCA