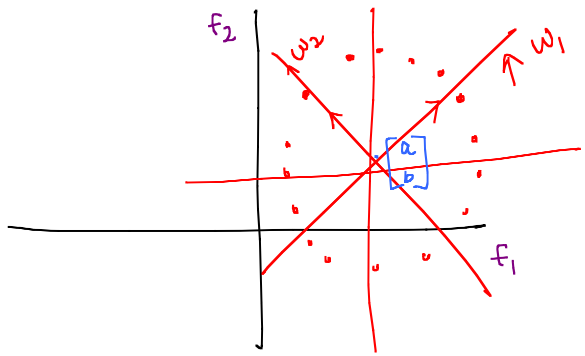
Issue 2 → Non-linear relationships



What would PCA give? $\{w_1, w_2\}$ and both important.

Relation between features

$$(f_1 - a)^2 + (f_2 - b)^2 = r^2$$

$$\Rightarrow \boxed{f_1^2 + a^2 - 2f_1 a + f_2^2 + b^2 - 2f_2 b - r^2 = 0} \leftarrow (*)$$

$x \in \mathbb{R}^2$ $\xrightarrow{\phi}$ $\phi(x)$

$$\underbrace{\begin{bmatrix} f_1 & f_2 \end{bmatrix}}_{\mathbb{R}^2} \longrightarrow \underbrace{\begin{bmatrix} 1 & f_1^2 & f_2^2 & f_1 f_2 & f_1 & f_2 \end{bmatrix}}_{\mathbb{R}^6}$$

Let $u \in \mathbb{R}^6$ $\underline{\begin{bmatrix} a^2+b^2-r^2 & 1 & 1 & 0 & -2a & -2b \end{bmatrix}}$

Each datapoint satisfies $\boxed{\phi(x)^T u = 0}$ $\equiv$ (*) $\Rightarrow$ the datapoints lie in a LINEAR subspace of $\mathbb{R}^6$!

IDEA: Transform features from low dimension $\mathbb{R}^d$ to high dimension $\mathbb{R}^D$

$$\underset{\mathbb{R}^d}{x} \xrightarrow{\phi} \underset{\mathbb{R}^D}{\phi(x)}$$

$\boxed{\phi(x)^T\phi(x)}$ instead of $\phi(x)\phi(x)^T$

- We already know how to handle the case when $D \gg n$

ISSUE — $\phi(x) \in \mathbb{R}^D$ may be too hard to compute.

$[f_1 \quad f_2 \quad f_3 \quad f_4]$

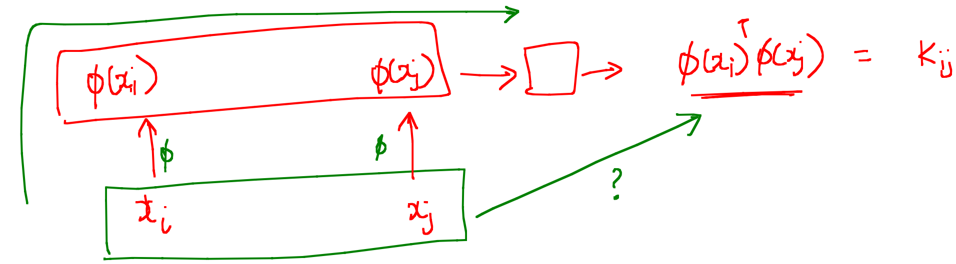
Cubic relations

$$\phi \Big(\quad [\underbrace{1}_{4C_0} \quad \underbrace{f_1 \quad f_2 \quad f_3 \quad f_4}_{4C_1} \quad \underbrace{f_1 f_2 \quad f_1 f_3 \quad \cdots \quad f_3 f_4}_{4C_2} \quad \underbrace{f_1 f_2 f_3 \quad f_1 f_2 f_4}_{4C_3} \cdots]$$

$1 + 4 + 6 + 4$

In general, d features, $\leq p$-th power

$$\sum_{i=0}^{p} {}^{d}C_i \approx O(d^p)$$



Example

$x = [f_1 \quad f_2]$ $\qquad$ $x' = [g_1 \quad g_2]$

Consider the function

$$(x^T x' + 1)^2 = \left([f_1 \quad f_2]\begin{bmatrix} g_1 \\ g_2 \end{bmatrix} + 1\right)^2 = (f_1 g_1 + f_2 g_2 + 1)^2$$

$$= f_1^2 g_1^2 + f_2^2 g_2^2 + 1 + 2 f_1 g_1 f_2 g_2 + 2 f_1 g_1 + 2 f_2 g_2$$

$$\boxed{(x^T x' + 1)^2}$$

$[f_1 \ f_2]$ $\quad$ $[g_1 \ g_2]$

$$= \underbrace{\begin{bmatrix} f_1^2 & f_2^2 & 1 & \sqrt{2} f_1 f_2 & \sqrt{2} f_1 & \sqrt{2} f_2 \end{bmatrix}} \begin{bmatrix} g_1^2 \\ g_2^2 \\ 1 \\ \sqrt{2} g_1 g_2 \\ \sqrt{2} g_1 \\ \sqrt{2} g_2 \end{bmatrix}$$

$$= \phi(x)^T \phi(x')$$

where

$$\phi(x) = \phi\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = \begin{bmatrix} a^2 \\ b^2 \\ 1 \\ \sqrt{2} ab \\ \sqrt{2} a \\ \sqrt{2} b \end{bmatrix}$$

INSIGHT:

$(x^T x' + 1)^2$ computes the dot-product in a "transformed space".

$[f_1 \quad f_2] \rightarrow [f_1^2 \quad f_2^2 \quad 1 \quad \sqrt{2} f_1 f_2 \quad \sqrt{2} f_1 \quad \sqrt{2} f_2]$

$[g_1 \quad g_2] \rightarrow [g_1^2 \quad g_2^2 \quad 1 \quad \sqrt{2} g_1 g_2 \quad \sqrt{2} g_1 \quad \sqrt{2} g_2]$

$$\boxed{\begin{array}{c} x \in \mathbb{R}^2 \\ \downarrow \\ \phi(x) \in \mathbb{R}^6 \end{array}}$$

WE MANAGED TO COMPUTE $\phi(x)^T \phi(x')$ WITHOUT EXPLICITLY COMPUTING $\phi(x)$

MORE EXAMPLES.

$x \in \mathbb{R}^d$:

$$k(x, x') = (x^T x' + 1)^p \qquad \text{for some } p \geq 1$$

→ can be shown to be a "valid" function

i.e., $\exists\, \phi : \mathbb{R}^d \to \mathbb{R}^D$ such that

$$k(x, x') = \phi(x)^T \phi(x')$$

EXERCISE: Compute the explicit ϕ for $p=3$ and $p=4$

ONE MORE EXAMPLE

$$k(x, x') = \exp\left(- \frac{\|x - x'\|^2}{2\sigma^2} \right) \quad \text{for some } \sigma > 0$$

RADIAL BASIS FUNCTION.

→ Can be shown to be a "Valid" map.

→ Interestingly, ϕ in this case maps x to an "infinite" dimensional space.

[Technicalities aside, can think of this as mapping a data point to a "function" and dot-products between functions become integrals].

KERNEL FUNCTION

Any function $k: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ which is a "valid" map is called a kernel function

$$k(x,x') = (x^T x' + 1)^p \Rightarrow \text{POLYNOMIAL KERNEL}$$

$$k(x,x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \Rightarrow \text{RADIAL BASIS / GAUSSIAN KERNEL}$$

Given a function $k: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, how can we say its a valid kernel?

METHOD 1: Exhibit a map ϕ explicity.

[might be hard sometimes]

METHOD 2: MERCER'S THEOREM (Informal)

$k: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a valid kernel

if and only if

$\phi(x)^T \phi(x)$

$\phi(x)\phi(x)^T$

(a) k is symmetric i.e., $k(x, x') = k(x', x)$.

(b) For any dataset $\{x_1, \ldots, x_n\}$, the matrix $K \in \mathbb{R}^{n \times n}$ where $K_{ij} = k(x_i, x_j)$ is POSITIVE SEMI DEFINITE

All e.values of K are non-negative. $\longrightarrow$

KERNEL PCA

- Input – $\{x_1, \dots, x_n\}$ $x_i \in \mathbb{R}^d$; kernel $k: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$.

- Step 1 : Compute $K \in \mathbb{R}^{n \times n}$ where

  $K_{ij} = k(x_i, x_j)$ $\forall i,j$

  $\longrightarrow$ • "Center the kernel"

- Step 2 : Compute $\beta_1, \dots, \beta_\ell$ eigen vectors & eigen values of K.

  $n\lambda_1 \geq \dots n\lambda_\ell$

  and normalize to get

  $\forall u$ $\alpha_u = \dfrac{\beta_u}{\sqrt{n\lambda_u}}$

X. Step 3 : $w_k = \phi(x)\alpha_k$ → Defeats the purpose because it needs $\phi(x)$.

- We cannot "reconstruct" the eigenvectors of the co-variance matrix.

- But we can still compute the "compressed" representation.

$$\phi(x_i)^T w_k = \phi(x_i)^T \left(\sum_{j=1}^{n} \phi(x_j)\, \alpha_{kj} \right)$$

$$= \sum_{j=1}^{n} \alpha_{kj}\, \phi(x_i)^T \phi(x_j) = \sum_{j=1}^{n} \alpha_{kj}\, K_{ij}$$

For downstream tasks, this is usually good enough.

- Modified Step 3 : Compute $\sum_{j=1}^{n} \alpha_{kj} K_{ij} \quad \forall k$

$$x_i \in \mathbb{R}^d \longrightarrow \left[\sum_{j=1}^{n} \alpha_{1j} K_{ij} , \sum_{j=1}^{n} \alpha_{2j} K_{ij} \cdots \sum_{j=1}^{n} \alpha_{\ell j} K_{ij} \right]$$

DETAILS : (Centering the kernel)

Given $K \in \mathbb{R}^{n \times n}$ $\quad K_{ij} = k(x_i, x_j) \quad \forall i,j$

$\{x_1, \ldots, x_n\}$

$\downarrow$

$\{\phi(x_1), \ldots, \phi(x_n)\}$

$\{\phi(x_1) - \mu, \ldots, \phi(x_n) - \mu\}$

$\mu = \frac{1}{n} \sum_i \phi(x_i)$

$(\phi(x_i) - \mu)^T (\phi(x_j) - \mu)$

$\boxed{\phi(x_i)^T \phi(x_j)}$

$\left. \begin{array}{l} -\mu^T \phi(x_j) \\ -\mu^T \phi(x_i) \\ +\mu^2 \end{array} \right\}$

Create a new kernel $K^c \leftarrow$ centered.

$$K^c_{ij} = K_{ij} - \theta_i \mathbb{1}_j^T - \mathbb{1}_i \theta_j^T + P \, \mathbb{1}_i \mathbb{1}_j^T$$

where

$$\left. \begin{array}{l} \theta_i = \frac{1}{n} \sum_{k=1}^{n} K_{ik} \quad \forall i \\ P = \frac{1}{n^2} \sum_{i,j} K_{ij} \end{array} \right\}$$