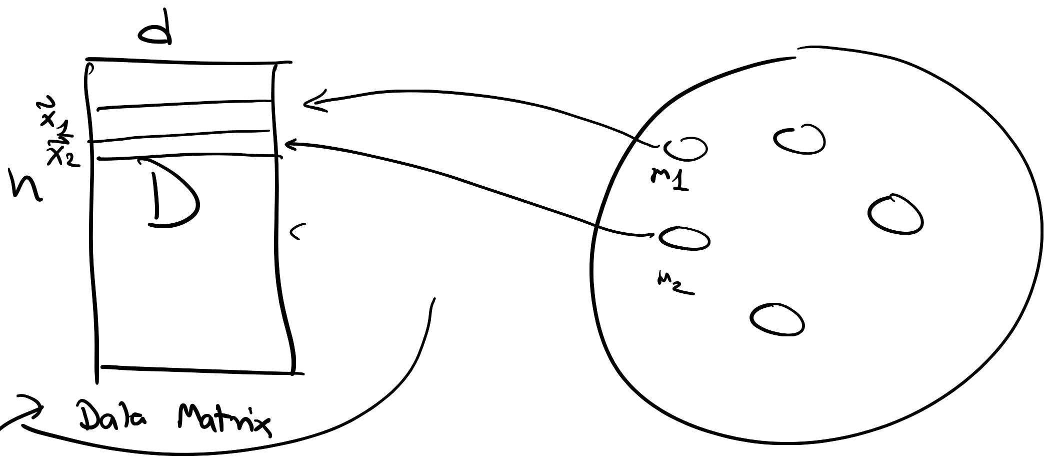


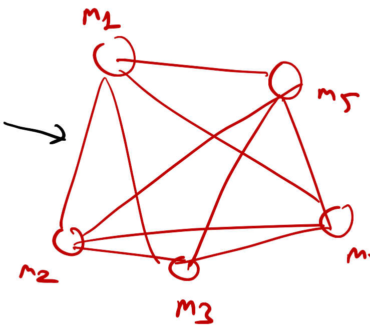
Data Sample

$O(nd)$

Similarity value $= k(m_1, m_2)$

Kernel function

Symmetric PSD kernel



Similarity matrix

Kernel matrix $\underline{K}$

$O(n^2)$

PSD

Positive Semi-Definite

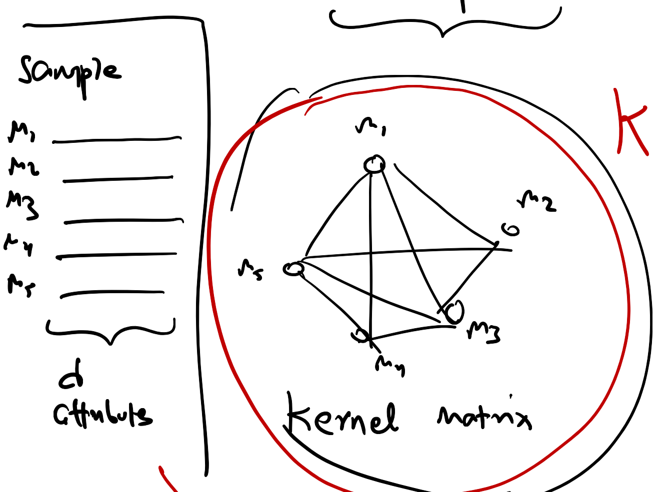
$$K(m_1, m_2) = k(m_2, m_1)$$

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

$$\vec{u}_1 \quad \vec{u}_2 \quad \cdots \quad \vec{u}_n$$

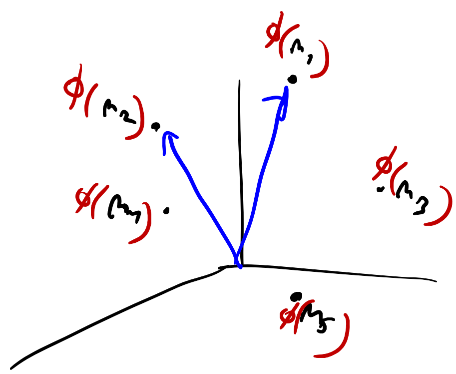
non-negative real eigenvalue

If k is a psd kernel, then $k(\vec{x}_i, \vec{x}_j)$ is a $\underbrace{\text{dot-product}}$ in $\underline{\text{some}}$ abstract vector space



Input space

non-linear transformation ϕ



$\mathbb{R}^\infty$

feature space

Conceptual

$$\boxed{k(m_1, m_2) = \phi(m_1)^T \phi(m_2)}$$

we never actually map $m_i \longrightarrow \phi(m_i)$

Can be computed in input space!

D

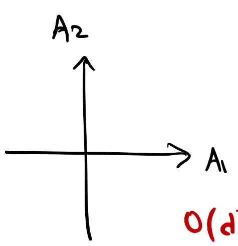
	A_1	A_2
$\vec{x}$	x_1	x_2

$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

$\xRightarrow{\phi}$ homogeneous quadratic

$\phi(D) =$

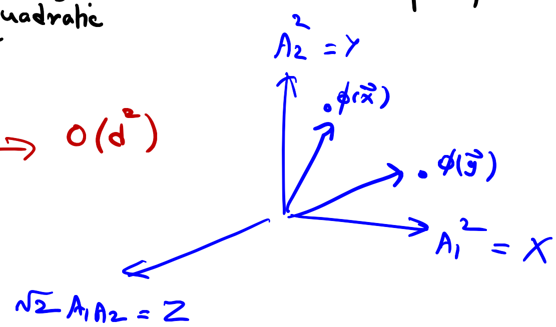
A_1^2	A_2^2	$A_1 A_2 (\sqrt{2})$
x_1^2	x_2^2	$\sqrt{2}\, x_1 x_2$



$\vec{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ d

$\phi(\vec{x}) = \begin{pmatrix} 1^2 \\ 2^2 \\ 1.2 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \\ 2\sqrt{2} \end{pmatrix}$

$O(d) \xrightarrow{\phi} O(d^2)$

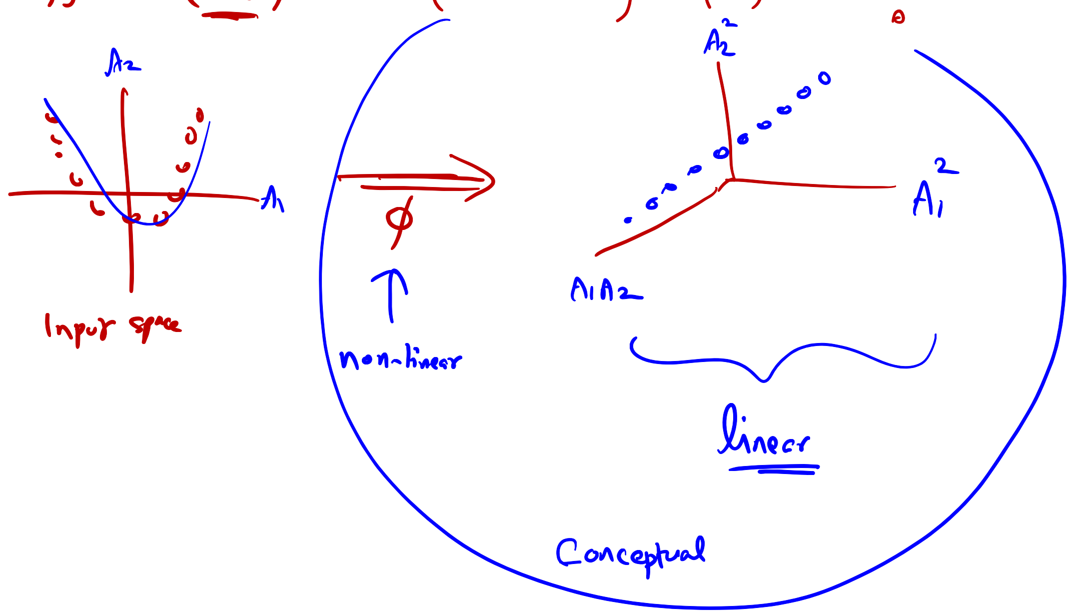


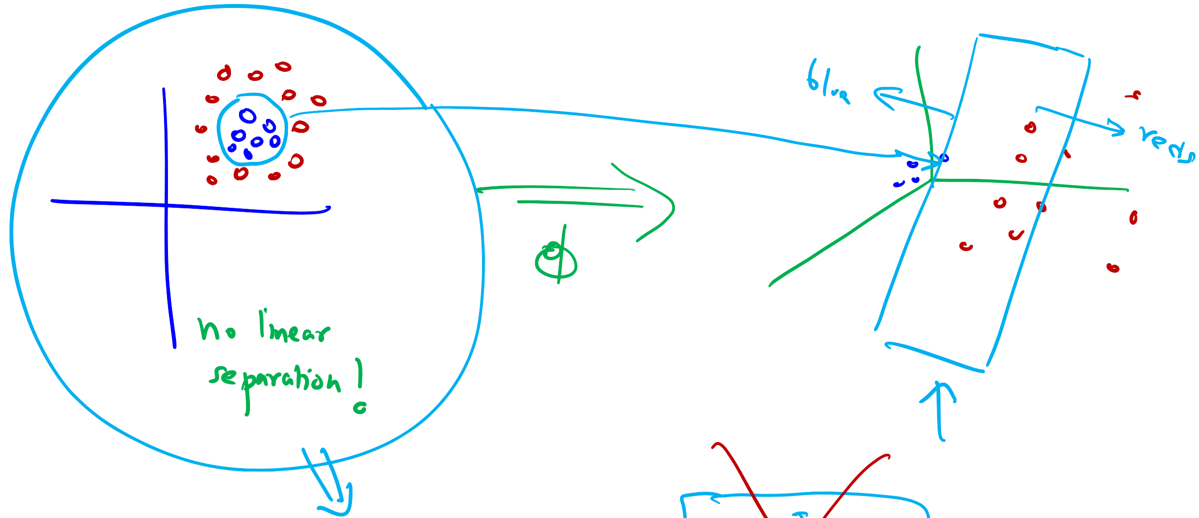
$\vec{y} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ $\quad \phi(\vec{y}) = \begin{pmatrix} 4 \\ 1 \\ 2\sqrt{2} \end{pmatrix}$

$$\phi(\vec{x})^T \phi(\vec{y}) = 4 + 4 + 8 = 16 = k(\vec{x}, \vec{y})$$

I want to avoid ϕ completely!

$$k(\vec{x}, \vec{y}) = (\underline{\vec{x}^T \vec{y}})^2 = (2 + 2)^2 = (4)^2 = 16 \,!$$





$$K = k(\vec{x}_i, \vec{x}_j) = \boxed{\cancel{\phi(\vec{x}_i)^T \phi(\vec{x}_j)}}$$

in same space!

ϕ inhomogeneous

$(q=2)$ degree

$c = 1$ constant

$$\vec{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$\phi(\vec{x}) = \begin{pmatrix} \text{constant} = c \\ \sqrt{2} x_1 \\ \sqrt{2} x_2 \\ x_1^2 \\ x_2^2 \\ \sqrt{2} x_1 x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ \sqrt{2} \\ 2\sqrt{2} \\ 1 \\ 4 \\ 2\sqrt{2} \end{pmatrix}$$

$$\boxed{k(\vec{x}, \vec{y}) = (\vec{x}^T \vec{y} + c)^2} = \phi(\vec{x})^T \phi(\vec{y})$$

Polynomial kernel (q : degree, c : constant)

$$k(\vec{x}, \vec{y}) = (\underbrace{\vec{x}^T \vec{y}}_{\text{scalar}} + c)^q$$

$\rightarrow$ $c = 0$: homogeneous

$\rightarrow$ $c > 0$: inhomogeneous

If $\vec{x} \in \mathbb{R}^d$, then $\phi(\vec{x}) \in \mathbb{R}^{\binom{d+q}{q}}$ space

$\approx O(d^q)$

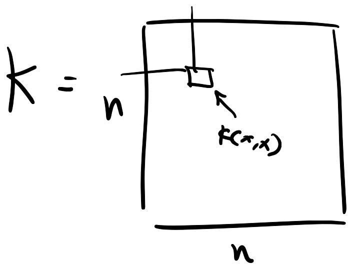
Gaussian kernel

$\mathbb{R}^d \rightarrow \mathbb{R}^\infty$

$\vec{x} \rightarrow \phi(\vec{x})$ impossible to physically construct $\phi(\vec{x})$

$$k(\vec{x}, \vec{y} \mid \sigma^2) = e^{-\left(\frac{\|\vec{x}-\vec{y}\|^2}{2\sigma^2}\right)}$$

you choose



linear kernel

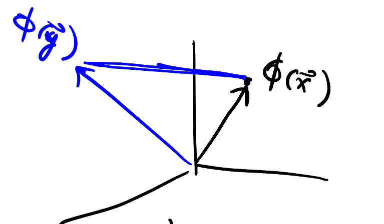
$q=1$ $\quad \underline{\underline{\vec{x}^T\vec{y}}} = k(\vec{x},\vec{y})$

kernel operations

1) length or norm of a point

$$k(\vec{x},\vec{x}) = \|\phi(\vec{x})\|^2$$

$$\sqrt{k(\vec{x},\vec{x})} = \|\phi(\vec{x})\|$$

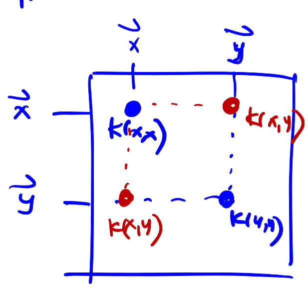


$\|\phi(x)\|$?

$$\underbrace{\phi(x)^T\phi(x)}_{k(\vec{x},\vec{x})} = \|\phi(x)\|^2$$

2) distance between $\phi(\vec{x})$ & $\phi(\vec{y})$ using only K

$$\|\phi(\vec{x}) - \phi(\vec{y})\|^2 = k(\vec{x},\vec{x}) + k(\vec{y},\vec{y}) - 2k(\vec{x},\vec{y})$$



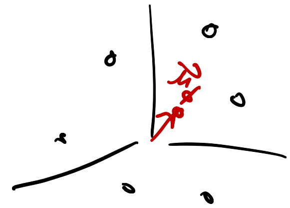
$$\|\phi(\vec{x}) - \phi(\vec{y})\|^2 =$$

$$(\phi(\vec{x}) - \phi(\vec{y}))^T(\phi(\vec{x}) - \phi(\vec{y}))$$

$$\phi(x)^T\phi(x) - \phi(x)^T\phi(y) - \phi(x)^T\phi(y) + \phi(y)^T\phi(y)$$

3) Mean vector in feature space via K ?

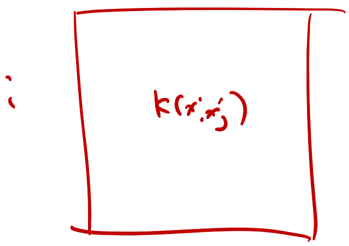
NOT POSSIBLE



$$\vec{\mu}_\phi = \frac{1}{n}\sum_{i=1}^{n}\phi(\vec{x}_i)$$

however we can compute

$$\boxed{\|\mu_\phi\|^2 = \frac{1}{n^2}\sum_i\sum_j k(\vec{x}_i,\vec{x}_j)}$$



average kernel/sim value

$$\|\vec{\mu}_\phi\|^2 = \mu_\phi^T\mu_\phi$$

$$= \frac{1}{n^2}\sum_i\sum_j \phi(\vec{x}_i)^T\phi(\vec{x}_j)$$

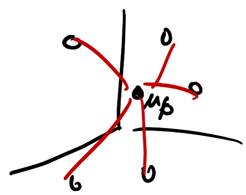
$$= \frac{1}{n^2}\sum_i\sum_j k(x_i,x_j)$$

---

4) Total variance via K

$$\text{totvar in feature space} = \underbrace{\frac{1}{n}\sum_{i=1}^{n}k(x_i,x_i)}_{\text{avg sq norm of a point in feature space}} - \underbrace{\frac{1}{n^2}\sum_i\sum_j k(x_i,x_j)}_{\text{sq norm of mean vector}}$$

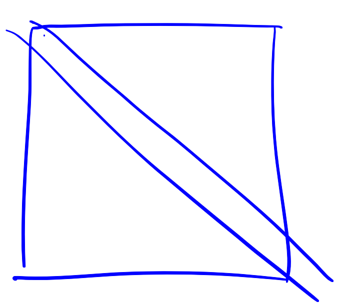
$$\text{avg } \|\phi(x_i)\|^2 - \|\mu_\phi\|^2$$



$$\text{totvar} = \frac{1}{n}\sum\|\phi(x_i) - \mu_\phi\|^2$$

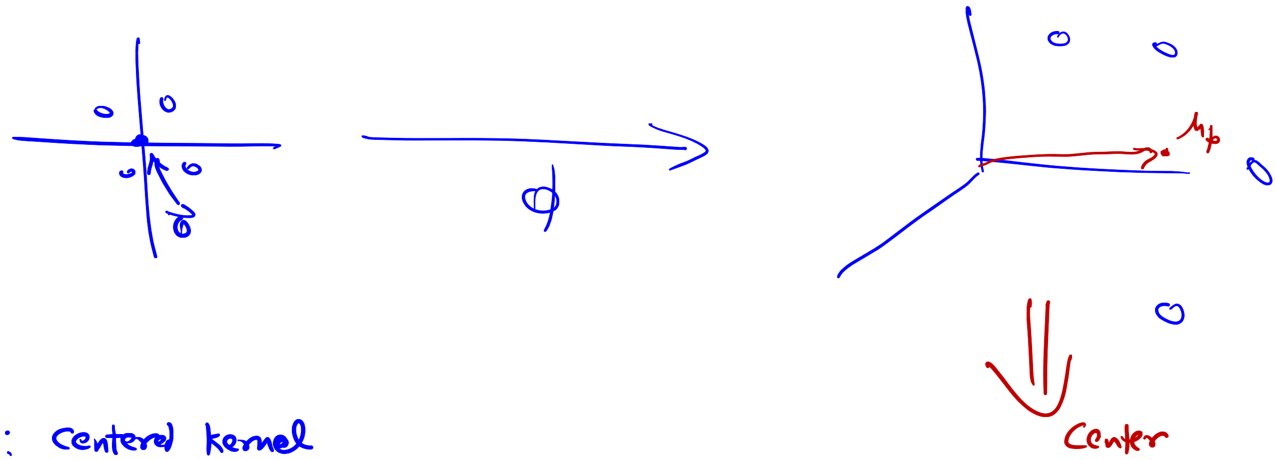
$$= \frac{1}{n}\sum\left(\phi(x_i)-\mu_\phi\right)^T\left(\phi(x_i)-\mu_\phi\right)$$

$$\mathrm{Var}(X) = E[X^2] - \left(E[X]\right)^2$$



avg diagonal − avg of matrix

5) Centering 6) angle 7) projections

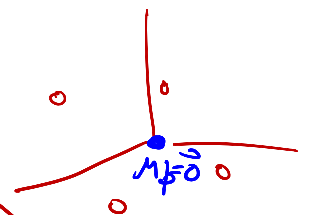


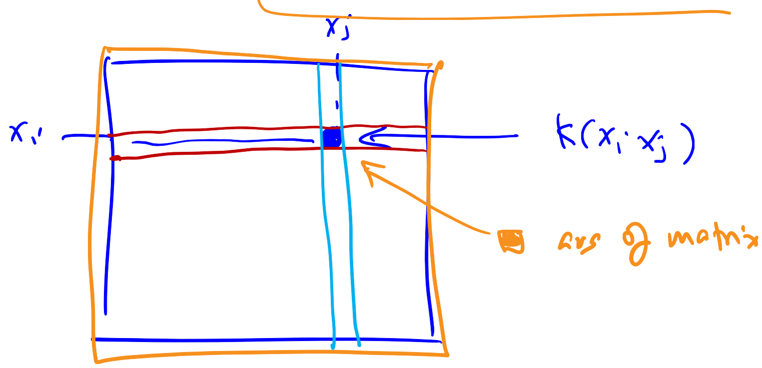
$\bar{K}$: centered kernel

$$\underbrace{\bar{K}(x_i, x_j)}_{\text{centered value}} = k(x_i, x_j) - \frac{1}{n}\sum_a k(x_i, x_a) - \frac{1}{n}\sum_b k(x_b, x_j) + \frac{1}{n^2}\sum_a \sum_b k(x_a, x_b)$$

avg sim of $\vec{x}_i$

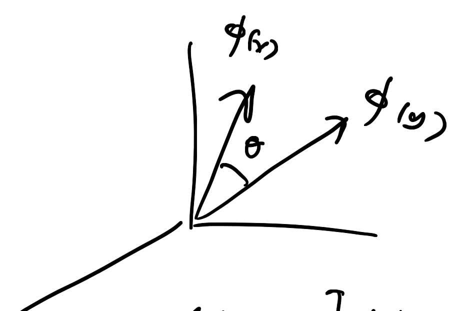
avg sim of $\vec{x}_j$





6) angle

$$\cos\theta = \frac{k(x,y)}{\sqrt{k(x,x) \cdot k(y,y)}}$$



$$\cos\theta = \left(\frac{\phi(x)}{\|\phi(x)\|}\right)^T \left(\frac{\phi(y)}{\|\phi(y)\|}\right)$$

7) projections

$$= \frac{k(x,y)}{\sqrt{k(x,x)}\,\sqrt{k(y,y)}}$$

handicap: we cannot construct a vector in feature space

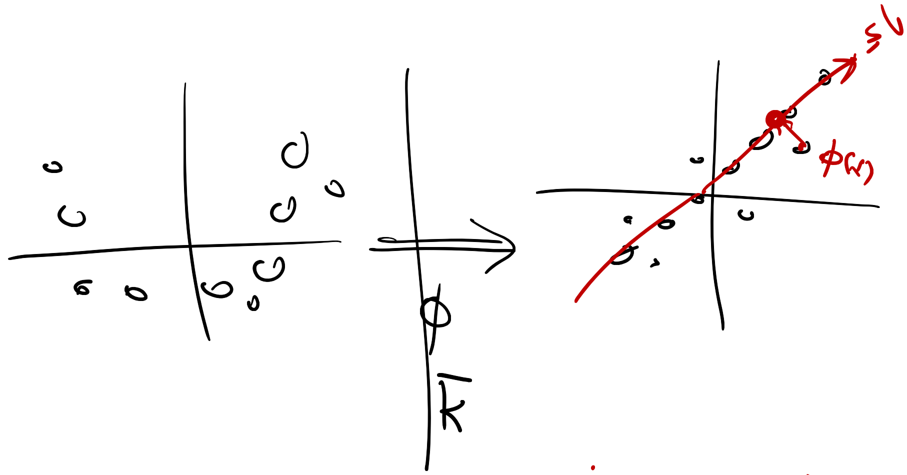
$\vec{\mu}_\phi = ?$

$\vec{u}_1 = ?$

$\vec{a}$ proj onto $\vec{u}_1$

$$\underbrace{\text{proj}_{\vec{u}_1}(\vec{a})}_{\checkmark} \cdot \cancel{\vec{u}_1}$$

$$\frac{\vec{u}_1^T \vec{a}}{\vec{u}_1^T \vec{u}_1}$$



$$\text{proj}_{\vec{u}_1}(\phi(x_i))$$

$$\boxed{\frac{\vec{u}_1^T \phi(x_i)}{\vec{u}_1^T \vec{u}_1} = \frac{c_1 \sum_a k(x_i, x_a)}{c_2 \sum_a \sum_b k(x_a, x_b)}}$$