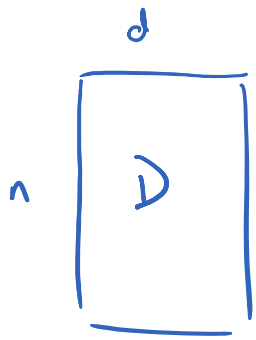
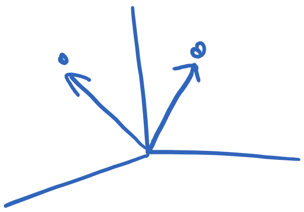
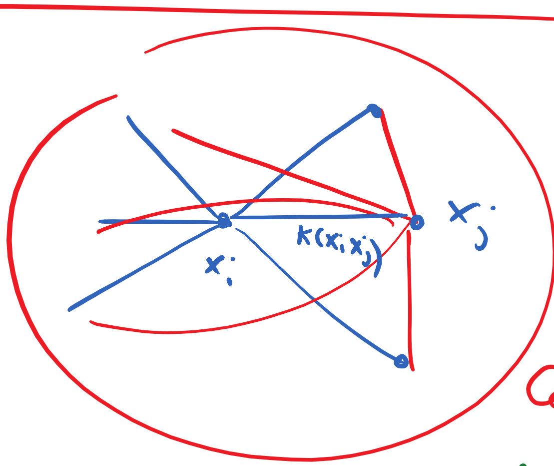


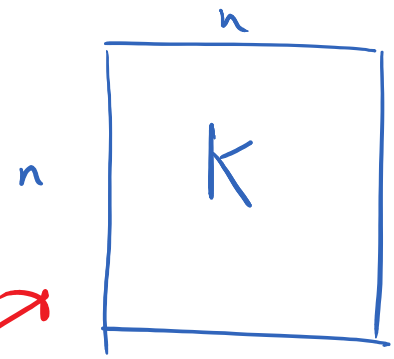
$\longrightarrow$ d- dimensions

$\longrightarrow$ change the basis set, PCA!



$K(x_i, x_j)$



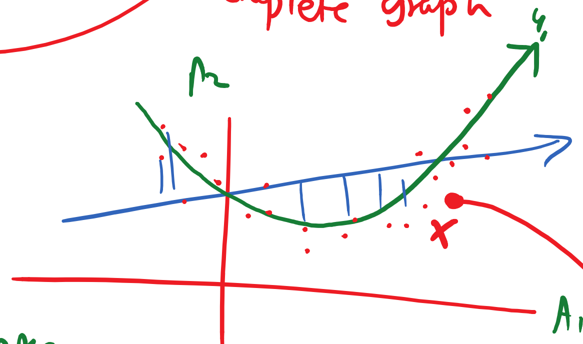


kernel matrix

Complete graph

	A_1	A_2
x_1	1	2
x_2	3	1
⋮		

Input space



ϕ

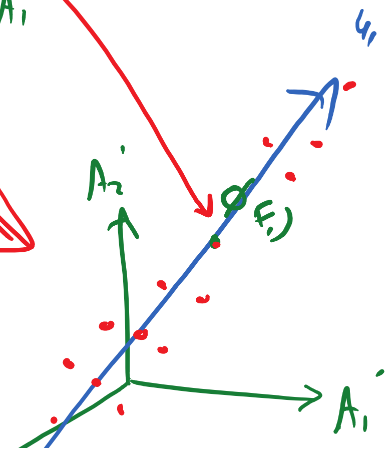
ϕ transformation

quadratic

homogeneous

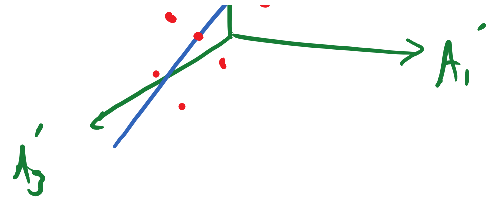
	$A_1' = A_1^2$	$A_2' = A_2^2$	$A_3' = A_1 A_2$
$\phi(x_1)$	1	4	2
$\phi(x_2)$	9	1	3

ϕ



$\phi(x_2)$	9	1	3

feature space



$u_1 = c_1 A_1' + c_2 A_2' + c_3 A_3'$

$u_1 = c_1 A_1^2 + c_2 A_2^2 + c_3 A_1 A_2$

Inhomogeneous quadratic

power up to 2

$\mathbb{R}^2 \to \mathbb{R}^6$

1	A_1	A_2	A_1^2	A_2^2	$A_1 \cdot A_2$
1	1	2	1	4	2
1	3	1	9	1	3

$\mathbb{R}^d \longrightarrow \mathbb{R}^{\text{high-dim.}}$

$\boxed{x} \mapsto \phi \longrightarrow \boxed{\phi(x)}$

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

$\phi \to \underbrace{\phi(x) \in \mathbb{R}^\infty}$

Infinite-dimensional space

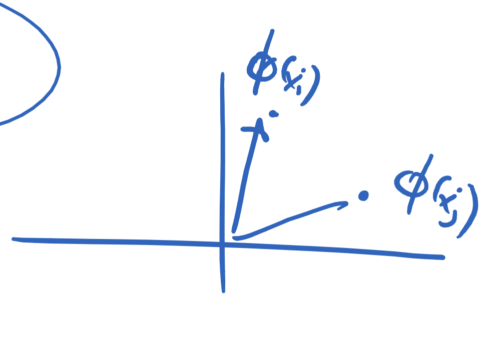
avoid the "physical" transformation

$\underbrace{K(x_i, x_j)}$ = dot product in feature space

nice function

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

has to be computable in input space



in ' input space

$k(x_i x_j)$: always dot product in **some** space

Condition on $K \in \mathbb{R}^{n\times n}$

1) symmetric

$$k(x_i, x_j) = k(x_j, x_i)$$

2) $K \leftarrow$ kernel matrix should be PSD
Positive semi-definite

→ eigenvalues of K are real & non-negative

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$$

$$\underbrace{u_1 \quad u_2 \quad \dots \quad u_n}_{\text{Orthogonal eigenvectors}}$$

Covariance!

Examples of kernel functions

⓪ linear kernel $k(x,y) = x^T y$

① Polynomial kernel of degree q

$$K(\vec{x},\vec{y}) = (\vec{x}^T\vec{y})^q$$

ϕ?

$\vec{x} \in \mathbb{R}^d$
$\vec{y} \in \mathbb{R}^d$

$q = 2$
$d = 2$

$$k(\vec{x},\vec{y}) = (x_1 y_1 + x_2 y_2)^2$$

$$= x_1^2 y_1^2 + x_2^2 y_2^2 + 2 x_1 y_1 x_2 y_2$$

$$= \begin{pmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2}\, x_1 x_2 \end{pmatrix}^T \begin{pmatrix} y_1^2 \\ y_2^2 \\ \sqrt{2}\, y_1 y_2 \end{pmatrix}$$

	A_1	A_2
$\vec{x}$	x_1	x_2
$\vec{y}$	y_1	y_2

	A_1^2	A_2^2	$\sqrt{2}A_1A_2$
$\phi(x)$	x_1^2	x_2^2	$\sqrt{2}\,x_1x_2$

	A_1	A_2	$\sqrt{2} A_1 A_2$
$\phi(x)$	x_1^2	x_2^2	$\sqrt{2} x_1 x_2$
$\phi(y)$	y_1^2	y_2^2	$\sqrt{2} y_1 y_2$

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

② Inhomogeneous polynomial

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

$q = 2$ $\quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ $\quad y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ $\quad d = 2$

$$(1 + \underbrace{x_1 y_1 + x_2 y_2}_{\Downarrow})^2$$

$$\underbrace{1^2 + 2(x_1 y_1 + x_2 y_2)} + \underbrace{(x_1 y_1 + x_2 y_2)^2}$$

$$1 + 2x_1 y_1 + 2 x_2 y_2 + \left(x_1^2 y_1^2 + x_2^2 y_2^2 + 2 x_1 y_1 x_2 y_2 \right)$$

$$\boxed{\phi(x) = \left(1, \sqrt{2} x_1, \sqrt{2} x_2, x_1^2, x_2^2, \sqrt{2} x_1 x_2\right)}$$

$$\underbrace{1}_{q=0} \quad \underbrace{\sqrt{2}\, A_1 \quad \sqrt{2}\, A_2}_{q=1} \quad \underbrace{A_1^2 \quad A_2^2 \quad \sqrt{2} A_1 A_2}_{q=2}$$

③ ~~exponential kernel~~

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

σ^2: spread parameter

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

dot product

ϕ ?

what is the feature space ?

Inversely proportional to distance

$$= \exp\left\{ -\gamma \underbrace{\|x - y\|^2} \right\} \qquad \gamma = \frac{1}{2\sigma^2}$$

$2\sigma^2$

what is the feature space?
$\mathbb{R}^\infty$

$$\|x-y\|^2 = \|x\|^2 + \|y\|^2 - 2x^Ty$$

$$e^{\|x-y\|^2} = \underbrace{e^{\|x\|^2} \cdot e^{\|y\|^2}} \cdot e^{-2x^Ty}$$

$$\exp(x) = \underbrace{1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots}_{\text{Infinite Sum}}$$

$$\propto \underbrace{\left[1 + (-2x^Ty) + \frac{(-2x^Ty)^2}{2!} + \dots\right]}$$

$$\underbrace{\phi(x) = (\, . \;\; . \;\; . \;\; . \;\; . \;\; . \,) \in \mathbb{R}^\infty}$$

---

④ Sigmoid 'kernel'

→ hyperbolic tangent

$$\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$

scalar

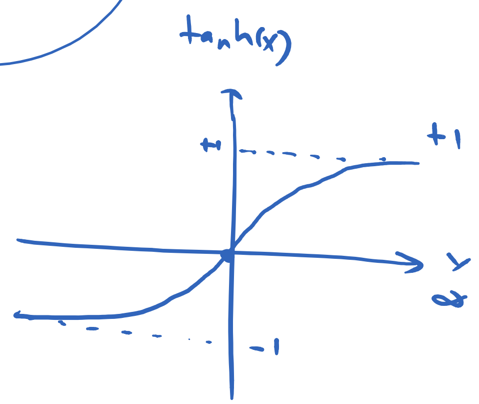
Common

a) $$\boxed{k(\vec{x}, \vec{y}) = \tanh(x^Ty)}$$

$$= \tanh\left(\sum_{i=1}^{d} x_i y_i\right)$$

not PSD

but "conditionally" PSD



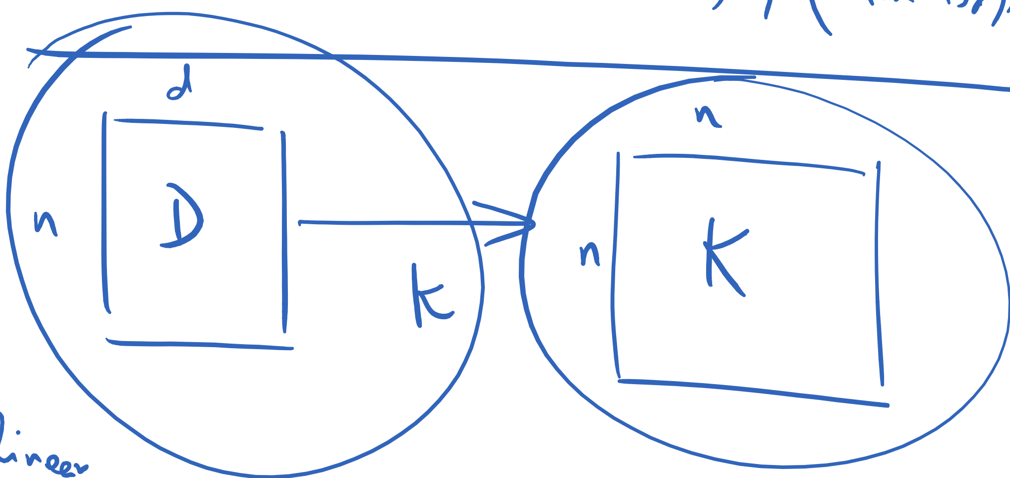
Mercer Sigmoid

b) $$k(\vec{x}, \vec{y}) = \sum_{i=1}^{d} \underbrace{\tanh(x_i)} \cdot \underbrace{\tanh(y_i)}$$

$$\Rightarrow \quad K(x,y) = \sum_{i=1} \tanh(x_i)\cdot\tanh(y_i)$$

PSD

$$= \begin{pmatrix} \tanh(x_1) \\ \tanh(x_2) \\ \vdots \\ \tanh(x_d) \end{pmatrix}^T \begin{pmatrix} \tanh(y_1) \\ \vdots \\ \tanh(y_d) \end{pmatrix}$$



linear only

throw away

$d << n$

$\underline{\underline{n^2}}$

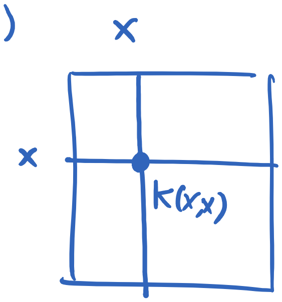
Non-linear as well

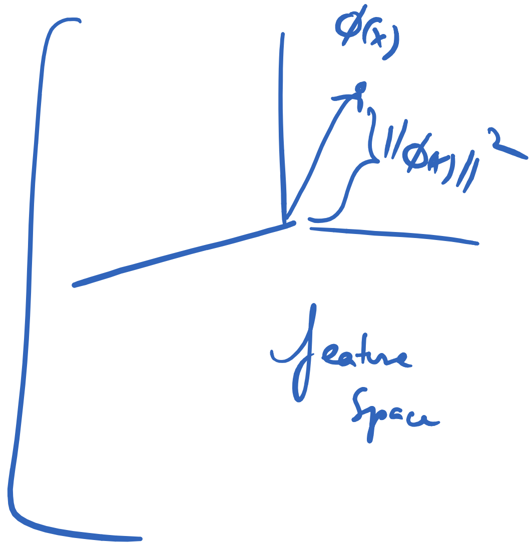
$K \in \mathbb{R}^{n\times n}$, PSD matrix

① length

$$\|\phi(x)\|^2 = \phi(x)^T\phi(x)$$
$$= \underline{K(x,x)}$$

squared length of $\phi(x)$

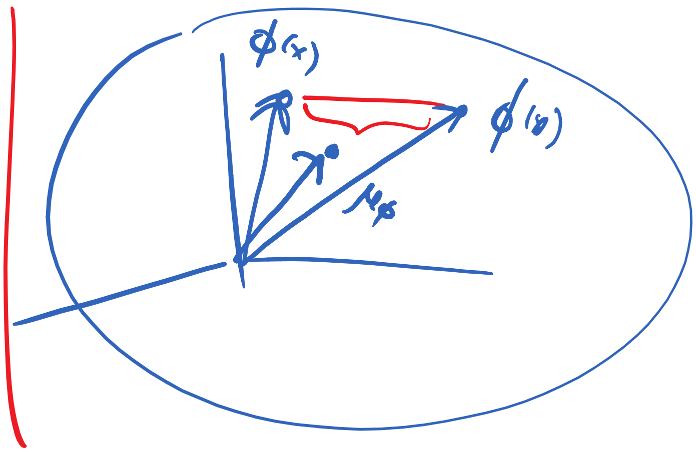


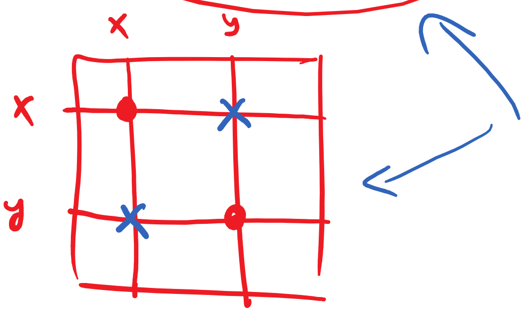


② distance in feature space

$$\|\phi(x)-\phi(y)\|^2 = \|\phi(x)\|^2 + \|\phi(y)\|^2 - 2\,\phi(x)^T\phi(y)$$

$$= k(x,x) + k(y,y) - 2k(x,y)$$





③ center of feature space

$$\mu_\phi = \frac{1}{n}\sum_{i=1}^{n}\phi(x_i)$$ ⟵ Cannot get the vector

$$\|\mu_\phi\|^2 = \mu_\phi^T\mu_\phi = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\phi(x_i)^T\phi(x_j)$$

$$\boxed{= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}k(x_i,x_j)}$$

↓

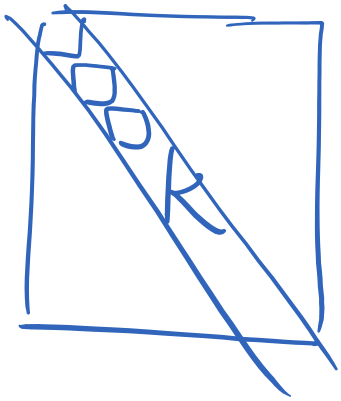
avg. kernel value

④ tot variance

$$\mathrm{totvar}(\phi(D)) = \frac{1}{n}\sum_{i=1}^{n}\|\phi(x_i)-\mu_\phi\|^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}k(x_i,x_i) - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}k(x_i,x_j)$$

$$= \underbrace{\frac{1}{n}\sum_{i=1} K(x_i, x_i)}_{\text{avg diagonal}} - \underbrace{\frac{1}{n^2}\sum_{i=1}\sum_{j=1} K(x_i, x_j)}_{\text{avg of matrix}}$$



(5) Compute angles in feature space

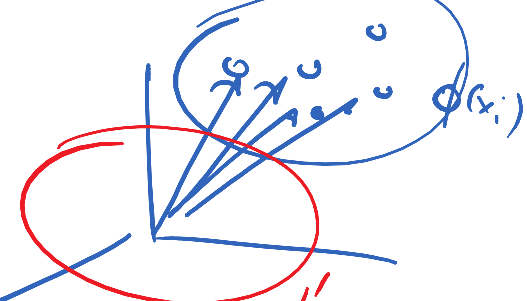
angle:

$$\cos\theta = \frac{K(x,y)}{\sqrt{K(x,x)}\,\sqrt{K(y,y)}} \qquad \frac{\phi(x)^T\,\phi(y)}{\|\phi(x)\|\;\|\phi(y)\|}$$

Correlation between "points" in feature

$$\sqrt{\|\phi(x)\|^2} \qquad K(x,x)$$

(6) Center data

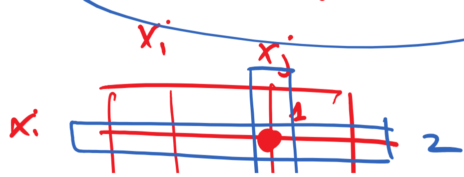


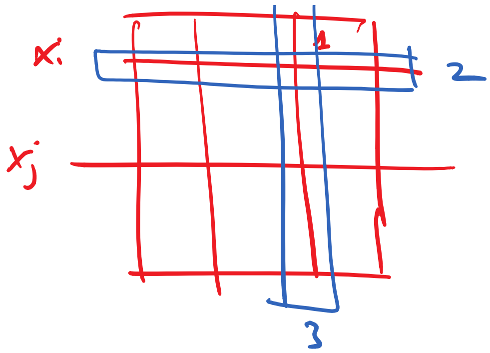
$$\underline{\underline{K'(x_i,x_j)}} = \underbrace{K(x_i,x_j)}_{1} \underbrace{- \frac{1}{n}\sum_a K(x_i,x_a)}_{2} \underbrace{- \frac{1}{n}\sum_a K(x_j,x_a)}_{3} + \frac{1}{n^2}\sum_a\sum_b K(x_a,x_b)$$

centered kernel

$$\phi' = \underline{\underline{\phi(x_i) - \mu_\phi}}$$

$K' = K$ after the μ shift





PCA over feature space

$K \leftarrow$ given

all operations here to be on K only!

Σ_ϕ is the cov matrix in F

$$\boxed{\Sigma_\phi u_1 = \lambda_1 u_1}$$

Compute dominant eigenvector of Σ_ϕ

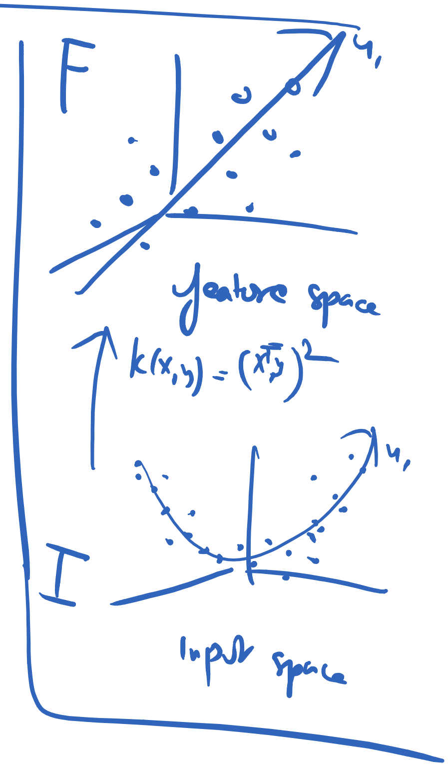
$\vec{u}_1$: in F

$$\boxed{\vec{u}_1 = \sum_{i=1}^{n} c_i \phi(x_i)}$$

$$\Sigma_\phi = \frac{1}{n} \sum_{i=1}^{n} \phi(x_i)\phi(x_i)^T$$

$$\boxed{K\vec{c} = n \cdot \lambda_1 \vec{c}} \longleftarrow \quad \vec{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ \vdots \\ c_n \end{pmatrix}$$

$\vec{c}$: the dominant eigenvector of the kernel matrix!



$\vec{u}_1$: unknown ? $= \sum_{i=1}^{n} c_i \phi(x_i)$

$a_1 = \phi(x_j)^T \vec{u}_1 = \phi(x_j)^T \left(\sum_{i=1}^{n} c_i \phi(x_i) \right)$

Projected point

$$a_1 = \sum_{i=1}^{n} c_i \, k(x_i, x_j)$$