A kernel function $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ must satisfy the *Mercer's condition*, which verifies that for any sequence of data points $x_1, \dots, x_n \in \mathbb{R}^d$ and coefficients $c_1, \dots, c_n \in \mathbb{R}$ the inequality

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_i\, c_j\, k(x_i, x_j) \geq 0$$

is satisfied. If it is the case, the kernel is called a *Mercer kernel*.

Conversely, the *representer theorem* states that if k is a Mercer kernel on $\mathbb{R}^d$, then there exists a Hilbert space (i.e., a finite or infinite dimensional $\mathbb{R}$-vector space with norm and scalar product) $\mathcal{F}$, the so-called feature space, and a continuous map $\varphi : \mathbb{R}^d \to \mathcal{F}$, such that

$$k(x, x') = \langle\, \varphi(x)\,,\, \varphi(x')\,\rangle_{\mathcal{F}} \quad \text{for all } x, x' \in \mathbb{R}^d.$$

Exercise 1: Mercer Kernels (3 × 20 P)

(a) *Show* that the following are Mercer kernels.

i. $k(x, x') = \langle x, x' \rangle$

$$\begin{aligned}\sum_i\sum_j c_i c_j k(x_i, x_j) &= \sum_i\sum_j c_i c_j \langle x_i, x_j\rangle \\ &= \sum_i\sum_j \langle c_i \cdot x_i, c_j \cdot x_j\rangle \\ &= \Big\langle \sum_i c_i x_i, \sum_j c_j x_j \Big\rangle \\ &= \Big\| \sum_i c_i x_i \Big\|^2 \geq 0\end{aligned}$$

Bilinearity:

$\alpha \cdot \langle x, y\rangle = \langle \alpha \cdot x, y\rangle = \langle x, \alpha \cdot y\rangle$

$\langle x, y\rangle + \langle x, z\rangle = \langle x, y+z\rangle$

$\varphi(x) = x$ $\qquad K = \Phi(X)\Phi(X)^T$

$K_{ij} = \langle x_i, x_j\rangle \to c^T K c \geq 0$

$= \langle \Phi(x_i), \Phi(x_j)\rangle$

ii. $k(x, x') = f(x) \cdot f(x')$ where $f : \mathbb{R}^d \to \mathbb{R}$ is an arbitrary continuous function

$$\begin{aligned}\sum_i\sum_j c_i c_j k(x_i, x_j) &= \sum_i\sum_j c_i c_j f(x_i)\cdot f(x_j) \\ &= \Big(\sum_i c_i f(x_i)\Big)\cdot\Big(\sum_j c_j f(x_j)\Big) \\ &= \Big(\sum_i c_i f(x_i)\Big)^2 \geq 0\end{aligned}$$

(b) Let k_1, k_2 be two Mercer kernels, for which we assume the existence of a finite-dimensional feature map associated to them. *Show* that the following are again Mercer kernels.

i. $k(x, x') = k_1(x, x') + k_2(x, x')$

$$\begin{aligned}\sum_i\sum_j c_i c_j k(x_i, x_j) &= \sum_i\sum_j c_i c_j \big(k_1(x_i, x_j) + k_2(x_i, x_j)\big) \\ &= \underbrace{\sum_i\sum_j c_i c_j k_1(x_i, x_j)}_{\geq 0} + \underbrace{\sum_i\sum_j c_i c_j k_2(x_i, x_j)}_{\geq 0} \geq 0\end{aligned}$$

ii. $k(x, x') = k_1(x, x') \cdot k_2(x, x')$

$$\begin{aligned}\sum_i\sum_j c_i c_j k(x_i, x_j) &= \sum_i\sum_j c_i c_j k_1(x_i, x_j)\cdot k_2(x_i, x_j) \\ &= \sum_i\sum_j c_i c_j \langle \Phi(x_i), \Phi(x_j)\rangle \langle \Psi(x_i), \Psi(x_j)\rangle \\ &= \sum_i\sum_j c_i c_j \Big(\sum_{m=1}^{d_1} \Phi_m(x_i)\Phi_m(x_j)\Big)\Big(\sum_{n=1}^{d_2} \Psi_n(x_i)\Psi_n(x_j)\Big) \\ &= \sum_m\sum_n\sum_i\sum_j c_i c_j \Phi_m(x_i)\Phi_m(x_j)\Psi_n(x_i)\Psi_n(x_j) \\ &= \sum_m\sum_n \Big(\sum_i c_i \Phi_m(x_i)\Psi_n(x_i)\Big)\Big(\sum_j c_j \Phi_m(x_j)\Psi_n(x_j)\Big) \\ &= \sum_m\sum_n \Big(\sum_i c_i \Phi_m(x_i)\Psi_n(x_i)\Big)^2 \geq 0\end{aligned}$$

$\langle x, y\rangle = \sum_{i=1} x_i y_i$

(c) *Show* using the results above that the polynomial kernel of degree d, where $k(x, x') = (\langle x, x'\rangle + \vartheta)^d$ and $\vartheta \in \mathbb{R}^+$, is a Mercer kernel.

$K^*_{ij} = \langle x_i, x_j\rangle$

$\Theta_{ij} = \vartheta$

$$c^T(K + \Theta)c = \underbrace{c^T K c}_{\geq 0} + \underbrace{c^T \Theta c}_{\geq 0} \geq 0$$

$$\underbrace{\underbrace{\underbrace{(\langle x, x'\rangle + \vartheta)}_{\text{kernel}}\cdot\underbrace{(\langle x, x'\rangle + \vartheta)}_{\text{kernel}}}_{\text{kernel}} \cdots}_{\text{kernel}}$$

Exercise 2: The Feature Map (4 × 10 P)

Consider the homogenous polynomial kernel k of degree 2 which is $k : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$, where

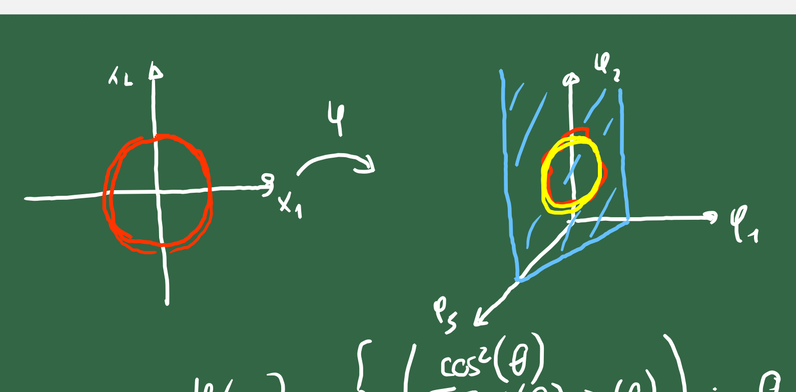
$$k(x, y) = \langle x, y\rangle^2 = \Big(\sum_{i=1}^{2} x_i\, y_i\Big)^2.$$

(a) *Show* that $\mathcal{F} = \mathbb{R}^3$ and $\varphi\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1^2 \\ \sqrt{2}\, x_1\, x_2 \\ x_2^2 \end{pmatrix}$ are possible choices for feature space and feature map.

$$\begin{aligned}\langle x, y\rangle^2 &= (x_1 y_1 + x_2 y_2)^2 \\ &= x_1^2 y_1^2 + 2 x_1 y_1 x_2 y_2 + x_2^2 y_2^2 \\ &= x_1^2 y_1^2 + \sqrt{2}\, x_1 x_2 \cdot \sqrt{2}\, y_1 y_2 + x_2^2 y_2^2 \\ &= \underbrace{\begin{pmatrix} x_1^2 \\ \sqrt{2}\, x_1 x_2 \\ x_2^2 \end{pmatrix}}_{\varphi(x)} \cdot \underbrace{\begin{pmatrix} y_1^2 \\ \sqrt{2}\, y_1 y_2 \\ y_2^2 \end{pmatrix}}_{\varphi(y)}\end{aligned}$$

$$\Psi(x) = \begin{bmatrix} x_1^2 \\ x_1 x_2 \\ x_1 x_2 \\ x_2^2 \end{bmatrix}$$

(b) Consider the unit circle $C = \left\{ \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \,;\, 0 \leq \theta < 2\pi \right\}$. *Show* that the image $\varphi(C)$ lies on a plane H in $\mathbb{R}^3$.



$$\begin{aligned}\varphi(C) &= \left\{ \begin{pmatrix} \cos^2(\theta) \\ \sqrt{2}\cos(\theta)\sin(\theta) \\ \sin^2(\theta) \end{pmatrix} ; 0 \leq \theta < 2\pi \right\} \\ &= \left\{ \begin{pmatrix} \cos^2(\theta) \\ \sqrt{2}\cos(\theta)\sin(\theta) \\ 1 - \cos^2(\theta) \end{pmatrix} ; 0 \leq \theta < 2\pi \right\} \\ &\subseteq \left\{ \begin{pmatrix} t \\ s \\ 1 - t \end{pmatrix} ; s, t \in \mathbb{R} \right\} = \operatorname{span}\left(\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\end{aligned}$$

(c) Consider the plane $A = \left\{ \begin{pmatrix} t \\ s \end{pmatrix} \,;\, t, s \in \mathbb{R} \right\}$. *Find* a point P in $\mathcal{F}$ which is not contained in $\varphi(A)$.

$$\begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} \notin \begin{pmatrix} t^2 \\ \sqrt{2}\, s t \\ s^2 \end{pmatrix} = \varphi(A)$$

(d) *Find* a feature map associated to the kernel $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ with $k(x, y) = \langle x, y\rangle^2 = \Big(\sum_{i=1}^{d} x_i\, y_i\Big)^2$.

$\Phi(x) = \big[(x_i^2)_i, (\sqrt{2}\, x_i x_j)_{i<j} \big] \in \mathbb{R}^{d\cdot(d+1)/2}$ $\quad\to$

$\Phi(x) = \big[(x_i x_j)_{ij} \big] \in \mathbb{R}^{d^2}$

$\to \langle \Phi(x), \Phi(y)\rangle \sim \mathcal{O}(d^2)$

$\to \langle x, y\rangle^2 \sim \mathcal{O}(d)$