Kernel methods

Eder Arley León Gómez

Universidad Nacional de Colombia

October 4, 2023

Schedule

1 Review

2 Reproduction Kernel Hilbert Spaces

3 Reproduction Kernel Kreĭm Spaces

4 Random Fourier Features

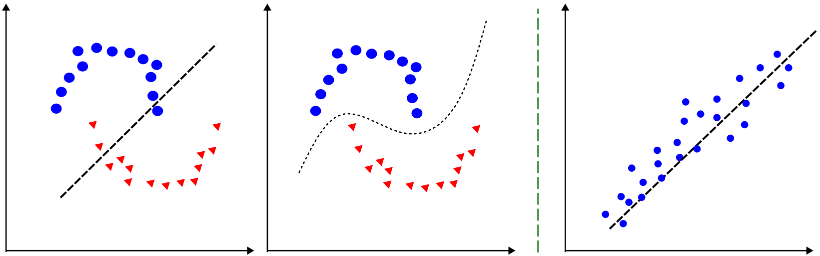
Review

Motivation

$$\{(\mathbf{x}_n, \mathbf{y}_n)\}_{n=1}^{N}, \text{ where } \mathbf{x}_n \in \mathbf{X} \text{ and } \mathbf{y}_n \in \mathbf{y}$$

A linear model finds a hyperplane $\mathbf{W}$ such that the output is:

$$f(\mathbf{X}) = \mathbf{W}^\top \mathbf{X} \tag{1}$$



Basic definitions

Norm

Let $\mathcal{H}$ be a vector space over $\mathbb{R}$. A function $||.||_{\mathcal{H}} : \mathcal{H} \to [0, \infty)$ is sad to be a norm on $\mathbb{R}$ if:

1. $||\mathbf{x}||_{\mathcal{H}} = 0$ if and only if $\mathbf{x} = 0$
2. $||\lambda\mathbf{x}||_{\mathcal{H}} = |\alpha|||\mathbf{x}||_{\mathcal{H}}, \ \forall \ \lambda \in R, \ \forall \ \mathbf{x} \in \mathbb{R}$
3. $||\mathbf{x} + \mathbf{y}||_{\mathcal{H}} \leq ||\mathbf{x}||_{\mathcal{H}} + ||\mathbf{y}||_{\mathcal{H}}, \ \forall \ \mathbf{x}, \mathbf{y} \in \mathcal{H}$

Cauchy sequence

A sequence $\{\mathbf{x}_n\}_{n=1}^{\infty}$ of real number is said to be Cauchy sequence if for every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that if $m, n \geq N$ $||\mathbf{x}_m - \mathbf{x}_n||_{\mathcal{H}} < \epsilon$

Banach space

Let $\mathcal{H}$ be a vector space equipped with a norm $||.||$. We say that $\mathcal{H}$ is Bananch space with respect to $||.||$ if every Cauchy sequence in $\mathcal{H}$ converged to a vector $\mathbf{x} \in \mathcal{H}$

Inner product space

Let $\mathcal{H}$ be a vector space over $\mathbb{R}$. A function $\langle .,. \rangle_{\mathcal{H}} : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ is said to be an inner product on $\mathcal{H}$ if:

1. Bilinearity

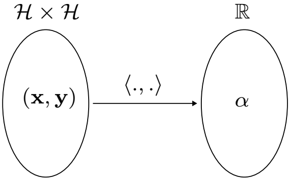
$$\langle \lambda_1 \mathbf{x} + \lambda_2 \mathbf{y}, \mathbf{z} \rangle_{\mathcal{H}} = \lambda_1 \langle \mathbf{x}, \mathbf{z} \rangle_{\mathcal{H}} + \lambda_2 \langle \mathbf{y}, \mathbf{z} \rangle_{\mathcal{H}}$$

2. Symmetry

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{H}} = \langle \mathbf{y}, \mathbf{x} \rangle_{\mathcal{H}}$$

3. Positive semi-definiteness

$\langle \mathbf{x}, \mathbf{x} \rangle_{\mathcal{H}} \geq 0$ and $\langle \mathbf{x}, \mathbf{x} \rangle_{\mathcal{H}} = 0$ if and only if $\mathbf{x} = 0$

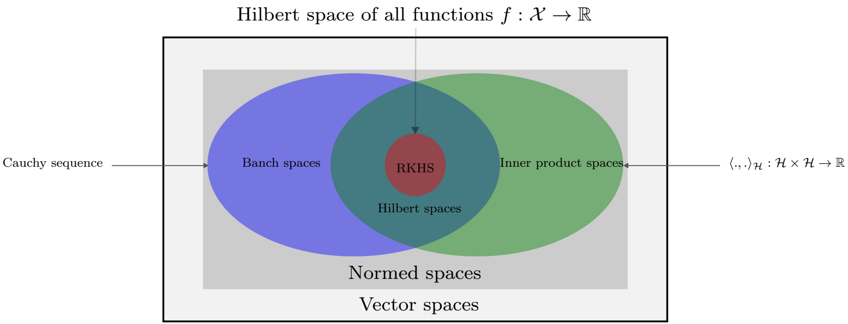


Hilbert spaces

Hilbert space

Hilbert space is a complete inner product space, i.e., it is a Banach space with an inner product.

Reproduction Kernel Hilbert Spaces



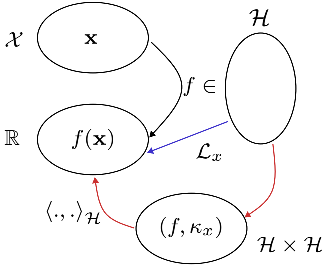
The term "reproducing" denotes the characteristic of the RKHS, enabling it to accurately represent specific functions by means of its **inner products**

Reproduction Kernel Hilbert Spaces

Definition of RKHS I

Let $\mathcal{X}$ be a set and $\mathcal{H}$ a Hilbert space of all functions $f : \mathcal{X} \to \mathbb{R}$. For each element $\mathbf{x} \in \mathcal{X}$, the evaluation functional is a linear functional that evaluates each $f \in H$ at the point $\mathbf{x}$, written:

$$\mathcal{L}_x : \mathcal{H} \to \mathbb{R}, \text{ where } \mathcal{L}_x(f) = f(\mathbf{x}) \text{ for all } f \in \mathcal{H}$$



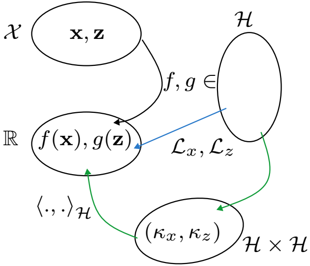
We say that $\mathcal{H}$ is a RKHS if, for all $\mathbf{x} \in \mathcal{X}$, $\mathcal{L}_x$ is continuous at every $f \in \mathcal{H}$.

Corollario: Reproducing property

$$\mathcal{L}_x(f) = f(\mathbf{x}) = \langle f, \kappa_x \rangle_{\mathcal{H}}$$

Definition of RKHS II

$$\mathcal{L}_x(f) = f(\mathbf{x}) = \langle f, \kappa_x \rangle_{\mathcal{H}} \qquad \mathcal{L}_z(g) = g(\mathbf{z}) = \langle g, \kappa_z \rangle_{\mathcal{H}}$$



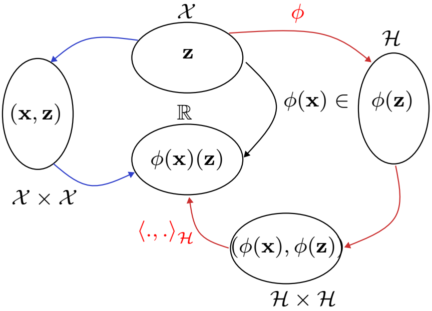
Kernel reproduction

Let $\mathcal{H}$ be a Hilbert space of $\mathcal{R}$-valued functions defined on a non-empty $\mathcal{X}$. A funtions $\kappa : \mathcal{H} \times \mathcal{H} \to \mathcal{R}$ is called Reproduction Kernel of $\mathcal{H}$ if it satisfies:

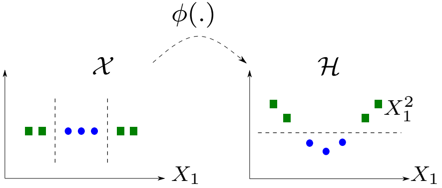
1. $\forall \mathbf{x} \in \mathcal{X}$, $\kappa(., \mathbf{x}) \in \mathcal{H}$
2. $\forall \mathbf{x} \in \mathcal{X}$, $\forall f \in \mathcal{H}$, $\langle f, \kappa(., \mathbf{x}) \rangle_{\mathcal{H}} = f(x)$ (the reproducing property)

Kernel trick

$$\kappa(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{z}) \rangle_{\mathcal{H}}$$



How to make comparisons?



Idea

$$\kappa(\mathbf{x},\mathbf{c}) = \left(1+\mathbf{x}^\top\mathbf{c}\right); \mathbf{x} = [\mathbf{x}_1,\mathbf{x}_2], \mathbf{x} = [\mathbf{c}_1,\mathbf{c}_2]$$

$$\kappa(\mathbf{x},\mathbf{c}) = \left(1+\begin{pmatrix}\mathbf{x}_1 & \mathbf{x}_2\end{pmatrix}\begin{pmatrix}\mathbf{c}_1\\ \mathbf{c}_2\end{pmatrix}\right)^2 = (1+\mathbf{c}_1\mathbf{x}_1+\mathbf{c}_2\mathbf{x}_2)^2$$

$$\kappa(\mathbf{x},\mathbf{c}) = \mathbf{c}_1^2\mathbf{x}_1^2+\mathbf{c}_2^2\mathbf{x}_2^2+2\mathbf{c}_1\mathbf{x}_1+2\mathbf{c}_2\mathbf{x}_2+2\mathbf{c}_1\mathbf{c}_2\mathbf{x}_1\mathbf{x}_2+1$$

So,

$$\phi(\mathbf{x}) = [1,\mathbf{x}_1^2,\sqrt{2}\mathbf{x}_1\mathbf{x}_2,\mathbf{x}_2^2,\sqrt{2}\mathbf{x}_1,\sqrt{2}\mathbf{x}_2]^\top$$

$$\phi(\mathbf{c}) = [1,\mathbf{c}_1^2,\sqrt{2}\mathbf{c}_1\mathbf{c}_2,\mathbf{c}_2^2,\sqrt{2}\mathbf{c}_1,\sqrt{2}\mathbf{c}_2]^\top$$

Kernel methods problem

Kernel trick

$$\kappa(\mathbf{x},\mathbf{y}) =< \phi(\mathbf{x}), \phi(\mathbf{y}) > \tag{2}$$

Main idea

- Define a comparison function: $\mathbf{K} : \mathcal{X} \times \mathcal{X}$
- Represent a set of n data points $S = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$ by the $n \times n$ matrix:

$$[\mathbf{K}]_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j) \tag{3}$$

Remarks

- K is always an $n \times n$ matrix, whatever the nature of data.
- Poor scalability with respect to the dataset size (n2 to compute and store K)

Positive-Definite Kernel Functions

- Lineal Kernel

$$\kappa(\mathbf{x},\mathbf{y}) = \mathbf{x}^\top \mathbf{y}$$

- Polinomial Kernel

$$\kappa(\mathbf{x},\mathbf{y}) = (\gamma.\mathbf{x}^\top \mathbf{y} + r)^d$$

- Radial Basic Function Kernel

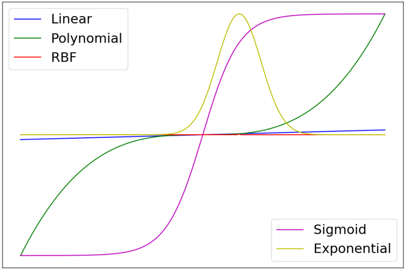
$$\kappa(\mathbf{x},\mathbf{y}) = exp\left(-\frac{||\mathbf{x}-\mathbf{y}||}{2\sigma^2}\right)$$

- Sigmoid Kernel

$$\kappa(\mathbf{x},\mathbf{y}) = tanh(\alpha.\mathbf{xy} + \beta)$$

- Exponential Kernel

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



Reproduction Kernel Kreĭm Spaces

Motivation

The essence of kernel methods resides in their utilization of a **positive definite linear function**, which can be associated with the **inner product** of two vectors in the **RKHS**.

Problem

The inherent and extrinsic attributes of real-world data pose challenges for positive-definite kernel functions.

Kreĭn spaces I

Kreĭn space is a generalization of the notion of Hilbert spaces, where the key difference is the fact that the inner products are indefinite.

Inner product positive

$$\langle \mathbf{x}, \mathbf{x} \rangle_{\mathcal{H}} \geq 0$$

Inner product negative

$$\langle \mathbf{x}, \mathbf{x} \rangle_{\mathcal{H}} \leq 0$$

Definition [1]

An inner product space $\mathcal{K}$ is a Kreĭn space is there exit two Hilbert spaces $\mathcal{H}_+$, $\mathcal{H}_-$ such that

- All $\mathbf{x} \in \mathcal{K}$ can be decomposed into $\mathbf{x} = \mathbf{x}_+ + \mathbf{x}_-$, where $\mathbf{x}_+ \in \mathcal{H}_+$ and $\mathbf{x}_- \in \mathcal{H}_-$
- $\forall \mathbf{x}, \mathbf{y} \in \mathcal{K}$, $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{K}} = \langle \mathbf{x}_+, \mathbf{y}_+ \rangle_{\mathcal{H}_+} - \langle \mathbf{x}_-, \mathbf{y}_- \rangle_{\mathcal{H}_-}$

[1] Kreĭn space can be defined on $\mathbb{R}$ or $\mathbb{C}$

Kreĭn space II

Associated Hilbert Space

Let $\mathcal{K}$ be a Kreĭn space with decomposition into Hilbert space $\mathcal{H}_+$ and $\mathcal{H}_-$. Then we denote by $\overline{\mathcal{K}}$ the associated Hilbert space defined by:

$$\overline{\mathcal{K}} = \mathcal{H}_+ + \mathcal{H}_-, \text{hence} < \mathbf{x}, \mathbf{y} >_{\overline{\mathcal{K}}} = < \mathbf{x}_+, \mathbf{y}_+ >_{\mathcal{H}_+} + < \mathbf{x}_-, \mathbf{y}_- >_{\mathcal{H}_-} \tag{4}$$

Likewise we can introduce the symbol $\ominus$ to indicate that:

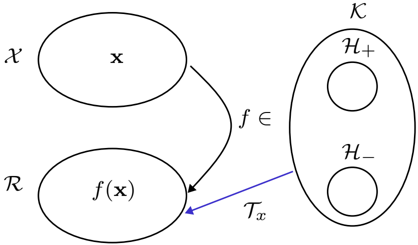
$$\mathcal{K} = \mathcal{H}_+ \ominus \mathcal{H}_- \tag{5}$$

The decomposition $\mathcal{K} = \mathcal{H}_+ \ominus \mathcal{H}_-$ is called the fundamental decomposition of the Kreın space $\mathcal{K}$

Definition of RKKS

A Kreĭn space $\mathcal{K}$ is a Reproducing Kreĭn Kernel Spaces (RKKS) if $\mathcal{K} \subset R^{\mathcal{X}}$ and the evaluation functional is continuous on $\mathcal{K}$ endowed with its strong topology

$$\mathcal{T}_x : \mathcal{K} \to \mathbb{R}, \text{ where } \mathcal{T}_x(f) = f(\mathbf{x}) \text{ for all } f \in \mathcal{K}$$



Reproducing property

$$\mathcal{T}_x = f(\mathbf{x}) = \langle f, \kappa_x \rangle_{\mathcal{K}}$$

From Kreĭn spaces to Kernels

Propositions

Let K be an RKKS with $\mathcal{K} = \mathcal{H}_+ \ominus \mathcal{H}_-$. Then:

- $\mathcal{H}_+$ and $\mathcal{H}_-$ are RKHS (with kernels κ_+ and κ_-)
- There is a unique symmetric $\kappa(\mathbf{x}, \mathbf{y})$ with $\kappa(\mathbf{x}, .) \in \mathcal{K}$ such that for all $f \in \mathcal{K}$, $< f, \kappa(\mathbf{x}, .) >_{\mathcal{K}} = f(\mathbf{x})$
- An indefinite kernel κ associated with a RKKS admits a positive decomposition $\kappa = \kappa_+ - \kappa_-$, with two positive kernels κ_+ and κ_-

Associated RKHS of RKKS

Let $\mathcal{K}$ be a RKKS with thje direct orthogonal sum decomposition into two RKHSs $\mathcal{H}_+$ and $\mathcal{H}_-$. Then the associated RKHS $\mathcal{K}$ endowed by $\mathcal{K}$ is defined with the positive inner product

$$< f, g >_{\mathcal{K}} = < f_+, g_+ >_{\mathcal{H}_+} + < f_-, g_- >_{\mathcal{H}_-}, \forall f, g \in \mathcal{K} \tag{6}$$

Indefinite kernels

- Epanechnikov kernel

$$\kappa(\mathbf{x},\mathbf{y}) = \left(1 - \frac{||\mathbf{x}-\mathbf{y}||^2}{\sigma}\right)^p, \text{for} \frac{||\mathbf{x}-\mathbf{y}||^2}{\sigma} \leqslant 1$$

- Gaussian Combination

$$\kappa(\mathbf{x},\mathbf{y}) = exp\left(-\frac{||\mathbf{x}-\mathbf{y}||^2}{\sigma_1}\right) + exp\left(-\frac{||\mathbf{x}-\mathbf{y}||^2}{\sigma_2}\right) + exp\left(-\frac{||\mathbf{x}-\mathbf{y}||^2}{\sigma_3}\right)$$

- Multiquadric kernel

$$\kappa(\mathbf{x},\mathbf{y}) = \sqrt{\frac{||\mathbf{x}-\mathbf{y}||^2}{\sigma} + c^2}$$

- Thin plate spline

$$\kappa(\mathbf{x},\mathbf{y}) = \frac{||\mathbf{x}-\mathbf{y}||^{2p}}{\sigma} In\left(\frac{||\mathbf{x}-\mathbf{y}||2}{\sigma}\right)$$

Random Fourier Features

Fourier transform

Idea

Fourier Transform is a mathematical model which helps to transform the signals between two different domains

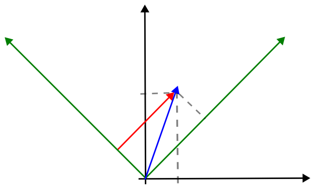
$$f(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kt + b_k \sin kt)$$

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos ktdt \qquad b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin ktdt$$

Remember

$$\langle f(x), g(x) \rangle = \int_a^b f(x) g(x) dx$$

$$|| \cos kx||^2 = || \sin kx||^2 = \pi$$



Wiener–Khinchin theorem

Power density

Power density refers to the amount of power in a signal per unit of bandwidth

$$S(f) = \lim_{T\to\infty} \frac{1}{T} \left| \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t)exp\left(-j2\pi ft\right) dt \right|^2$$

Wiener-khintchine relationships

The spectral density (or power density) $s(w)$ is definited as the Fourier transform of the correlation function, where:

$$S(w) = \int_{-\infty}^{\infty} R(\tau)exp\left(-jw\tau\right) d\tau \qquad R(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(w)exp\left(jw\tau\right) dw$$

The sufficient condition for the validity of the above equation is:

$$\int_{-\infty}^{\infty} |R(\tau)d\tau| < \infty$$

The Wiener-Khinchin relationships are a part of the Wiener-Khinchin theorem, which states that $R(\tau)$ is a positive definite correlation function if $S(w) \geq 0 \forall w$

Main idea

Basic idea

Let's go back to a lower-dimensional representation, using random Fourier features

Approximate the inner product $\kappa(\mathbf{x},\mathbf{y}) =< \phi(\mathbf{x}), \phi(\mathbf{y}) >_v$ with a random mapping $z : R^D \to R^R$ where $D << R$

$$\kappa(\mathbf{x},\mathbf{y}) =< \phi(\mathbf{x}), \phi(\mathbf{y}) >_v \approx z(\mathbf{x})^\top z(\mathbf{y}) \tag{7}$$

Where $z(.)$ is a good projection of ϕ

Translation invariant kernel

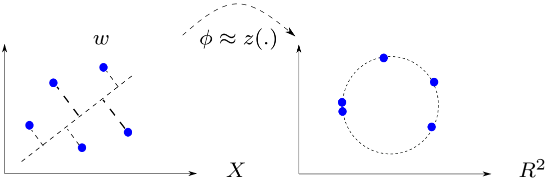
A Kernel $\kappa : \mathcal{H} \times \mathcal{H} \to \mathcal{R}$ is called translation invariant, if it only depends on the difference between its argument:

$$\forall_{\mathbf{x},\mathbf{y}} \in Z,\ \kappa(\mathbf{x}, \mathbf{y}) = \varphi(\mathbf{x} - \mathbf{y})$$

For some $\varphi : \mathcal{H} \to \mathcal{R}$. Such a function φ is called positive definite if the corresponding kernel φ is positive definite.

Random Fourier Features

"Each component of the feature map $z(\boldsymbol{x})$ projects $\boldsymbol{x}$ onto a random direction w drawn from the Fourier transform"



$z(\mathbf{x}) = \cos(w^\top \mathbf{x} + b)$; $w \in R^R$, $w \in R$

Bochner theorem

A continuous function $\varphi : \mathcal{H} \to \mathcal{R}$ is positive definite if and only if it is the Fourier transform of a symmetric and positive finite measure $\mu \in M(R^d)$.

$$\varphi(\mathbf{x} - \mathbf{y}) = \underbrace{\int p(w) exp\left(jw\Delta\right) dw} = E\left[\xi_w(\mathbf{x})\xi_w(\mathbf{y})^*\right] \tag{8}$$

Math review

$$\kappa(\mathbf{x},\mathbf{y}) = \kappa(\mathbf{x}-\mathbf{y}) = \int p(w) e^{jw(\mathbf{x}-\mathbf{y})} dw = E_w\left[\xi_w(\mathbf{x})\xi_w(\mathbf{y})^*\right] \tag{9}$$

where $\xi_w(\mathbf{x}) = e^{jw^\top \mathbf{x}}$

$$\kappa(\mathbf{x},\mathbf{y}) = \kappa(\mathbf{x}-\mathbf{y}) = E_w\left[e^{jw^\top \mathbf{x}} e^{-jw^\top \mathbf{y}}\right] = \int p(w)\left[\cos(w^\top(\mathbf{x}-\mathbf{y})) + j\sin(w^\top(\mathbf{x}-\mathbf{y}))\right]$$

$$\kappa(\mathbf{x},\mathbf{y}) = \kappa(\mathbf{x}-\mathbf{y}) = \int p(w)\cos(w^\top(\mathbf{x}-\mathbf{y})) + \underbrace{j\int p(w)\sin(w^\top(\mathbf{x}-\mathbf{y}))}_{0}$$

$$\hat{\kappa}(\mathbf{x},\mathbf{y}) = \hat{\kappa}(\mathbf{x}-\mathbf{y}) = E_w\left[z_w(\mathbf{x})z_w(\mathbf{y})\right]$$

$$z(\mathbf{x}) = \sqrt{\frac{2}{D}}\left[\cos(w_1^\top \mathbf{x} + b_1), \cos(w_2^\top \mathbf{x} + b_2), \ldots, \cos(w_D^\top \mathbf{x} + b_D)\right]$$

References

Dino Sejdinovic, Arthur Gretton
What is an RKHS?

Reproducing Kernel Hilbert Spaces Machine Learning
https://ngilshie.github.io/jekyll/update/2018/02/01/RKHS.html.

Rahimi, Ali and Recht, Benjamin
Random Features for Large-Scale Kernel Machines
Advances in neural information processing systems (2007).

Random Fourier Features
https://gregorygundersen.com/blog/2019/12/23/random-fourier-features/.

Thank you!