Gaussian processes, Signatures and Kernelizations

Josef Teichmann

ETH Zürich

March 2022

Example 1: randomized neural networks

By George Cybenko, Kurt Hornik et al. shallow neural networks

$$\Big\{\sum_i \alpha_i\varphi(\langle\beta_i, .\rangle + \gamma_i) \,|\, \alpha_i \in \mathbb{R}^n, \beta_i \in \mathbb{R}^d, \gamma_i \in \mathbb{R}\Big\}$$

are dense in $C([0,1]^d;\mathbb{R}^n)$. Finding parameters $\alpha_i, \beta_i, \gamma_i$ is a non-linear regression task, i.e. a generically non-convex optimization problem.

By Ali Rahimi, Benjamin Recht et al. it makes sense to consider β_i, γ_i randomly chosen according to certain distributions (together with the number of nodes) to return to a possibly convex optimization problem (depending on the loss function). This can then be related to kernel methods by considering the randomly chosen basis functions as approximations of a kernel eigensystem (compare to work of Nicholas Nelsen, Andrew Stuart). Neural tangent kernels take up this point of view and dynamize it again, see Arthur Jacot et al..

Approximation bounds and algorithmic feasibility often avoid the curse of dimension here and are an active area of research.

Example 2: signature methods

By Terry Lyons et al. linear functionals on signature of a (continuous) finite variation or rough path u form a point separating algebra of path space functionals on paths starting at 0. The zeroth component of the path is chosen time t here:

$$\Big\{ \sum_{k\geq 0, i_1,\ldots,i_k \in \{0,\ldots,d\}} \alpha_{i_1\cdots i_k} \int_{0\leq t_1 \leq \ldots t_k \leq t} du^{i_1}(t_1)\cdots du^{i_k}(t_k) \,|\alpha_{i_1,\ldots,i_k} \in \mathbb{R} \Big\}$$

Input space: space of finite variation path (extended by time), a rough path space of paths starting at 0.

Output space: for simplicity $\mathbb{R}$.

Example 2: signature methods

By Ilya Chevyrev, Harald Oberhauser et al. one can associate a kernelization of signature methods considering signature basis elements as an eigensystem of a kernel.

By Christa Cuchiero, Lukas Gonon et al. a randomized version of signature is given on path space (starting at 0) by

$$\Big\{\sum_j \alpha_j X_t^j(u) \,|\, \text{ where } dX_t = \sum_{i=0}^{d} \varphi(A_i X_t + b_i) du^i(t), X_0 \neq 0\Big\}$$

with A_i , b_i appropriately chosen random matrices according to certain distributions, and φ is an activation functions.

In contrast to the finite dimensional theory we have the following features: signature basis are generically *unbounded* on path spaces, signature itself does not depend on parameters over which one optimzes, i.e. it is solely a regression basis, whereas randomized signature is actually not signature and also unbounded.

Applications in Finance

- approximation of path space functionals, or more generally, predictable strategies by neural networks on relevant factors or signature basis on path space.
- Examples: deep hedging, deep portfolio optimization, deep drift estimation, signature based pricing and hedging, sig-SDEs, reservoir computing for learning dyncamics, stochastic optimization, stochastic games beyond Markovian paradigms, etc.
- some of these applications are quite successful, but still lack a full theoretical foundation why the non-convex optimization problem can be solved so efficiently or why existing approximation results are generically sufficient.

Applications in Finance

- approximation of path space functionals, or more generally, predictable strategies by neural networks on relevant factors or signature basis on path space.
- Examples: deep hedging, deep portfolio optimization, deep drift estimation, signature based pricing and hedging, sig-SDEs, reservoir computing for learning dyncamics, stochastic optimization, stochastic games beyond Markovian paradigms, etc.
- some of these applications are quite successful, but still lack a full theoretical foundation why the non-convex optimization problem can be solved so efficiently or why existing approximation results are generically sufficient.

Applications in Finance

- approximation of path space functionals, or more generally, predictable strategies by neural networks on relevant factors or signature basis on path space.
- Examples: deep hedging, deep portfolio optimization, deep drift estimation, signature based pricing and hedging, sig-SDEs, reservoir computing for learning dyncamics, stochastic optimization, stochastic games beyond Markovian paradigms, etc.
- some of these applications are quite successful, but still lack a full theoretical foundation why the non-convex optimization problem can be solved so efficiently or why existing approximation results are generically sufficient.

Goal of the talk

- develop a unified framework for approximations by signatures, neural networks, or combinations of it on finite or infinite dimensional spaces, on compact state spaces or beyond. This is important since in applications varieties of those input spaces appear.
- approximations beyond uniform or L^p norms should be included. This is important since regularizing procedures often lead to finer topologies.
- Randomization procedures should be applicable since this is an important ingredient in many algorithms, in particular kernelizations should be possible.

We shall work with compact spaces, weighted spaces, weak-$*$-topologized spaces as input spaces. Output spaces will be just Banach spaces.

Kernelizations

Let E be a topological space and H a separable Hilbert space. We are interested in maps $\varphi : E \to H$ such that $x \mapsto \langle l, \varphi(x) \rangle$, for $l \in E'$ lies dense in an appropriate function space $\mathcal{F}$.

Then one can define a positive definite kernel $k(x, y) : \langle \varphi(x), \varphi(y) \rangle$ for $x, y \in E$ and obtains that its reproducing kernel Hilbert space

$$H_k = \text{ closure of } \langle k_x := k(x, .) \text{ for } x \in E \rangle$$

with respect to the scalar product $\langle k_x, k_y \rangle_k := k(x, y)$ for $x, y \in E$, is densely embedded into $\mathcal{F}$ if and only if E is densely embedded into $\mathcal{F}$ (universality). Approximation of elements of $\mathcal{F}$ can be understood from the topology of H_k (representer theorems).

Gaussian processes

We can also consider a Gaussian process on E with covariance function k by looking at $x \mapsto \sum_i \langle e_i, \varphi(x) \rangle X_i$, where (X_i) is an i.i.d. sequence of standard normal random variables. The reproducing kernel Hilbert space H_k appears then as space of means $x \mapsto m(x)$ which appear through equivalent measure changes. Approximation of elements of $\mathcal{F}$ can be seen from a Bayesian perspective.

In both viewpoints it is of interest to consider spaces $\mathcal{F}$ of maps on E which are unbounded, since signature itself is.

Bernstein polynomials

A simple and beautiful application of the law of large numbers (LLN) is Sergey Bernstein's proof of Weierstrass approximation theorem:

A Bernstein polynomial of type (n, k) is defined by

$$B_{n,k}(x) = \binom{n}{k} x^k (1-x)^{n-k} \quad (k = 0, 1, \ldots, n)\,. \tag{1}$$

Then every continuous function f on $[0, 1]$ can be uniformly approximated by the following polynomial

$$B_n^f(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) B_{n,k}(x)\,,$$

where a quantitative estimate is given below.

Bernstein polynomials

Let (X_n) be a sequence of independent, identically distributed Bernoulli random variables with success parameter $x \in [0,1]$, then by LLN

$$\frac{X_1 + \ldots + X_n}{n} \to x$$

almost surely. We furthermore have

$$P\Big[X_1 + \ldots + X_n = k\Big] = B_{n,k}(x)\,.$$

Denote by S_n the sum $X_1 + \ldots + X_n$.

Bernstein polynomials

Whence

$$\begin{aligned}|B_n^f(x)-f(x)| &= \left|E\left[f\left(\frac{S_n}{n}\right)-f(x)\right]\right| \le E\left[\left|f\left(\frac{S_n}{n}\right)-f(x)\right|\right] \\ &\le 2\sup_u |f(u)|\ P\left[\left|\frac{S_n}{n}-x\right|>\delta\right] \\ &+ \sup_{|u-v|\le\delta}|f(u)-f(v)|\ P\left[\left|\frac{S_n}{n}-x\right|\le\delta\right].\end{aligned}$$

Since f is uniformly continuous we can bound the second term on the right hand side by ϵ for small enough δ. Due to Chebychev's inequality the first term is bounded by

$$2\sup_u |f(u)|\frac{x(1-x)}{n\delta^2} \le \frac{1}{2n\delta^2}\sup_u |f(u)| \le \epsilon\,,$$

for n large enough. Therefore

$$\|B_n^f - f\|_\infty \longrightarrow 0 \text{ for } n\to\infty.$$

C^k version of Bernstein approximation

A bit less know is that B_n^f also converges in C^k to f if $f \in C^k([0,1])$: we actually have to understand for this purpose, e.g., the case $k = 1$ that

$$f'(x) = \lim_{n\to\infty} E\Big[f'\big(\frac{S_n}{n}\big)\Big] = \lim_{n\to\infty} E\Big[f\big(\frac{S_n}{n}\big)\frac{B'_{n,S_n}(x)}{B_{n,S_n}(x)}\Big]$$

uniformly on $[0,1]$, which is a sort of integration by parts for $f \in C^1([0,1])$. The higher dimensional cases are analogous.

This can be seen as a discrete analogue of the famous formula $E[f^{(k)}(X)] = E[f(X)H_k(X)]$ for standard Gaussian random variables.

Weierstrass approximation theorem

This proves in particular the following theorem:

The polynomials are dense in $C([0,1]) = C([0,1],\mathbb{R})$ (Weierstrass approximation theorem).

A substantial generalization of this result tells that on compact topological Hausdorff spaces K every point separating subalgebra of the algebra of continuous functions $C(K) := C(K;\mathbb{R})$ is actually dense, too (Stone-Weierstrass approximation theorem). Point separating just means that for every two points $x \neq y$ there is a function $f \in A$ such that $f(x) \neq f(y)$.

There is an order theoretic version of this theorem and Bernstein's proof also paves the path towards a probabilistic version of this theorem.

Proof of the Stone Weierstrass approximation theorem

Let K be a compact topological Hausdorff space and let $A \subset C(K)$ be a point separating subalgebra ((sub-)algebras here always contain the 1). Let $f \in C(K)$ and $\epsilon > 0$ be fixed. Then we can proceed as follows:

- With $g \in A$, we have that $|g| \in \overline{A}$. Indeed $g(K) \subset [a, b]$ for some a, b, and take a polynomial p which approximates $x \mapsto |x|$ on $[a, b]$ up to accuracy ϵ. Then $\||g| - p(g)\|_\infty \leq \epsilon$, however $p(g) \in A$.
- With $g, h \in A$ we have that $\max(g, h) = \frac{|g+h|}{2} + \frac{|g-h|}{2} \in \overline{A}$.
- With $g, h \in \overline{A}$ we have that $\max(g, h) \in \overline{A}$.

Proof of the Stone Weierstrass approximation theorem

- For every $x \in K$ we construct $f_x \in \overline{A}$ such that $f_x \leq f + \epsilon$ and $f_x(x) = f(x)$. Indeed we can find (point separation) for every $z \in K$ a function $g_{x,z} \in A$ with $g_{x,z}(x) = f(x)$ and $g_{x,z}(z) = f(z)$. Then there exists an open neighborhood $V_z \ni z$ such that $g_{x,z}|_{V_z} \leq f|_{V_z} + \epsilon$. Due to compactness there is a finite subcover of (V_z) indexed by $z_1, \ldots, z_n \in K$. Define now $f_x = \min(g_{x,z_1}, \ldots, g_{x,z_n}) \in \overline{A}$.
- With an analogue argument we can construct an open cover (U_x) such that $f_x \geq f - \epsilon$ on $U_x \ni x$, which has again a finite subcover indexed by $x_1, \ldots, x_m$. Define now $g = \max(f_{x_1}, \ldots, f_{x_m}) \in \overline{A}$, then $f - \epsilon \leq g \leq f + \epsilon$ globally.

Remarks

- We could equally take a point separating, linear subspace A such that with $f, g \in \overline{A}$ also $\max(f, g) \in \overline{A}$ (order theoretic version of the Stone Weierstrass approximation theorem).
- A probabilistic version could look as follows: let ν be a measure with full support on K and let $\mu_{n,x} = g_{n,x}\nu$ be a family of probability measures converging weakly to δ_x as $n \to \infty$, for $x \in K$. Assume that $x \mapsto g_{n,x}(y)$ is continuous for every y in the support of ν. Then the span of $x \mapsto g_{n,x}(y)$ is dense in $C(K)$.

Vector valued Stone Weierstrass approximation theorem

Let Y be a Banach space. Let $B \subset C(K;Y)$ be an A-submodule, where A a point separating subalgebra of $C(K)$. Assume furthermore that $(g(x))_{g\in B}$ is a dense family in Y for every $x \in K$. Then B is dense in $C(K;Y)$.

The proof is simple: without restriction we can assume that $A = C(K)$ and that B is closed. Take $f \in C(K;Y)$ and choose $\epsilon > 0$. For every $x \in K$ choose $g_x \in B$ such that $g_x(x) = f(x)$. Then $(\{y \in K \mid \|f(y) - g_x(y)\| < \epsilon\})$ is an open cover of K which has a finite subcover indexed by $x_1, \ldots, x_n \in X$. Choose a partition of unity $\sum_i \psi_i = 1$ for this finite subcover, then $g := \sum_i \psi_i g_{x_i} \in B$ is approximating f up to accuracy ϵ.

Weighted Spaces

For several applications it is necessary to go beyond compact spaces. We therefore consider weighted spaces (E, ρ), i.e. topological Hausdorff spaces with $\rho : E \to \mathbb{R}_{\geq 1}$ such that $\{\rho \leq R\}$ is compact for all R, where a similar analysis as on compact spaces is possible.

We consider the closure $B^\rho(E)$ of bounded continuous functions $C_b(E;\mathbb{R}) = C_b(E)$ with respect to the ρ-norm

$$\|f\|_\rho := \sup_x \frac{|f(x)|}{\rho(x)} .$$

In a similar manner we can define $B^\rho(E; Y)$ for vector valued functions.

Stone Weierstrass approximation theorem for weighted spaces E

Let A a point separating subalgebra of $B^\rho(E)$ such that for a point separating subspace $\tilde{A} \subset A$ the function $\exp(|l|) \in B^\rho(E)$ for $l \in \tilde{A}$. Then A is dense in $B^\rho(E)$.

Assume first that A consists soley of bounded functions, then the additional condition is automatically satisfied. In this case the proof follows directly from the compact case: it is sufficient to show that $f \in C_b(E) \subset B^\rho(E)$ can be approximated by elements from A. Choose $R > 0$, then we can find $g \in A$, such that g is close to f on $\{\rho \leq R\}$ with distance less than $1 > \epsilon > 0$. Assume f has range bounded by M, whence there is a polynomial p which closely approximates on $[-M - \|g\|_\infty - 1, M + \|g\|_\infty + 1]$ a function being $x \mapsto x$ on $[-M - 1, M + 1]$ and bounded by $M + 1$ otherwise. Consequently $p(g) \in A$ is close to f with distance less than $\epsilon + \frac{M+1}{R}$, but now globally in ρ-norm (if R is chosen big enough such that M/R is small).

Stone Weierstrass approximation theorem for weighted spaces E

Assume now the general case: the additional condition means that by polynomial approximation $\sin(l)$ and $\cos(l)$ lie in the closure of A for $l \in \tilde{A}$, whence the *subalgebra*

$$\{\alpha_1 \sin(l_1) + \alpha_2 \cos(l_2) | \alpha_i \in \mathbb{R}, l_i \in \tilde{A}\}$$

of globally bounded functions lies in the closure of A.

By the first result we can conclude that A is dense.

Vector valued Stone Weierstrass approximation theorem for weighted spaces E

Let Y be a Banach space. Let $B \subset B^\rho(E;Y)$ be an A-submodule, where A a point separating subalgebra of $B^\rho(E)$ of bounded continuous functions (or under the previous more general condition). Assume furthermore that $(g(x))_{g \in B}$ is a dense family in Y for every $x \in E$. Then B is dense in $B^\rho(E;Y)$.

Again without restriction we can assume that $A = B^\rho(E;Y)$ and again it is sufficient to show that $f \in C_b(E;Y) \subset B^\rho(E;Y)$ can be approximated by elements from B. Choose $R > 0$, then we can choose $g \in B$, such that g is close to f on $\{\rho \leq R\}$ with distance less than $1/3 > \epsilon > 0$. Assume without restriction that f has range bounded by $1/3$. The function $h = 1 \wedge \frac{1}{5/3 + \|g\|^2}$ is bounded continuous on E, therefore it lies in A. $hg \in B$ is still close to f with distance less than $\epsilon + \frac{1}{R}$ but now globally (if R is chosen large enough as above).

Remark

- We can replace the Banach space Y by any locally convex vector space and obtain analogue results for the locally convex spaces of vector valued continuous functions on K or E, respectively.
- In the real valued case an order theoretic version is possible, too.

Nachbin type theorems

Leopoldo Nachbin proved versions of the Stone-Weierstrass theorem for the C^k topology, where the point separating subalgebra is subject to an additional condition, the so called Nachbin condition.

We do neither enter differentiability theory on infinite dimensional spaces nor the precise details on the Nachbin condition, but just take the following definition.

Let (E,ρ) be weighted space and A a Banach space and a point separating subspace of functions $A \subset B^\rho(E;\mathbb{R}^n)$ such that for all bounded $f_1,\ldots,f_r \in A$ and all C^k functions g it holds that $g(f_1,\ldots,f_r) \in A$. Furthermore for sequences of C^k functions converging in C^k, also the corresponding compositions converge in A (with respect to its Banach space topology). In such a case we call A a C^k algebra. (we could apply *convenient calculus* here and consider Lip^k spaces instead.

Example: $C^k(K;\mathbb{R}^m)$, where K is a ball in $\mathbb{R}^n$.

Nachbin type theorems

Let A be a C^k algebra and $B \subset A$ a subalgebra such that the set of all possible $g(f_1, \ldots, f_n)$ for $f_i \in B$ bounded and g a C^k function is dense. Then B is already dense in A.

The proof is a simple applications of polynomials being C^k dense in $C^k([0, 1])$.

In a similar manner vector valued version can be defined.

Universal approximation theorems (UAT)

this part seems important.

Universal approximation theorems aim for easy constructions of subalgebras or submodules on weighted spaces in order to apply Stone Weierstrass type approximation theorems.

We shall introduce the notion of *activating families* and *additive families* for this purpose.

Additive point separating families

Let E be a weighted space. A set of bounded continuous functions $\mathcal{L} \subset B^{\rho}(E)$ is called additive point separating family if it is closed under addition, contains 1, -1 and is point separating.

We remark that this definition also makes sense for vector valued functions.

Activating families

Let Y be a Banach space. A family Φ of continuous functions $\varphi : \mathbb{R} \to \mathbb{R}$ is called activating if the space

$$A_\Phi := \Big\{ \sum_i \alpha_i \varphi_i(\beta_i . + \gamma_i) + \alpha_0 | \text{ for } \alpha, \alpha_0 \in Y,\ \beta_i \in \mathbb{N}, \gamma \in \mathbb{Z}, \varphi_i \in \Phi, n \in \mathbb{N} \Big\}$$

is dense in $C([a, b]; Y)$ for any real $a < b$

Typically Φ is a singleton ('an activation function'). Notice that it is sufficient that this property holds for $Y = \mathbb{R}$, since then it holds for all finite dimensional spaces, whence for all finite dimensional subspaces of Y, wherefrom the general assertion follows by vector valued Stone-Weierstrass on $[a, b]$.

We call Φ a C^k activating family if the topology in the above statement is actually C^k.

UAT

Let Y be a Banach space, E a weighted space, Φ and activating family of functions and $\mathcal{L}$ an additive family, then

$$\mathrm{NN}_\Phi = \Big\{ \sum_i \alpha_i \varphi_i(l_i) + \alpha_0 | \text{ for } \alpha_i \in Y,\ l_i \in \mathcal{L} \text{ and } n \in \mathbb{N} \Big\}$$

is dense in $B^\rho(E; Y)$.

Proof of UAT

For the proof we have to show that the closure B of NN_Φ is a $B^\rho(E)$ submodule which satisfies the condition that $(g(x))_{g\in B}$ is dense for every $x \in E$.

Assume first that $Y = \mathbb{R}$, then the algebra A generated by $\mathcal{L}$ is point separating and therefore dense. This algebra, however, lies in the closure of NN_Φ. Indeed consider $l \in \mathcal{L}$, then $\sin(l)$ as well as $\cos(l)$ lie in the closure since we can approximate *sin* and *cos* by functions from A_Φ uniformly (notice that l has bounded range). Therefore $\sin(k_1 l_1 + \cdots + k_n l_n)$ and $\cos(k_1 l_1 + \cdots + k_n l_n)$ lie in the closure, for $l_i \in \mathcal{L}$ and $k_i \in \mathbb{N}$ (addivity!). By uniform trigonometric approximation we obtain therefore that all polynomials of elements from $\mathcal{L}$ lie in the closure, whence we can conclude by real valued Stone-Weierstrass.

For the general case it is sufficient to show it for finite dimensional subspaces of Y, where it clearly holds.

Remark

- We could replace $[a, b]$ in the definition of activating families above a weighted locally convex vector space Z such that $\exp(|l|) \in B^\rho(Z)$ for each $l \in Z'$. Consider now activating families taking values in Z, then an analogous result holds true.
- We can also consider activating families of functions $\varphi : Z \to Z$, α_i should then be linear maps from Z to Y.
- Elements of NN_Φ are called *neural networks with activating family* Φ *initialized by* $\mathcal{L}$.
- The space of real valued neural networks NN_Φ is again an additive family. Whence deeper networks are dense, too.

Activating families

- If $\varphi : \mathbb{R} \to \mathbb{R}$ is a discriminatory function, i.e. a Borel measure μ is vanishing if and only if

$$\int_a^b \varphi(\beta x \pm \gamma)\mu(dx) = 0$$

for all real numbers $\beta, \gamma \in \mathbb{R}$, then $\Phi = \{\varphi\}$ is an activating family.

- If $\varphi : \mathbb{R} \to \mathbb{R}$ is bounded and non-constant, then it is discriminatory. The same result holds with respect to C^k-topologies (see work of Kurt Hornik).
- If $\varphi(x) = \max(0, x)$, then $\Phi = \{\varphi\}$ is an activating family.

UAT for C^k algebras

Let A be a C^k algebra on a weighted space E, Φ a C^k activating family of functions and $\mathcal{L}$ an additive family in A such that $g(l_1, \dots, l_r)$ for all possible C^k functions g and $l_1, \dots, l_r \in L$ are dense in A, then

$$\mathrm{NN}_\Phi = \left\{ \sum_i \alpha_i \varphi_i \circ l_i \middle| \text{ for } \alpha_i \in \mathbb{R}^n,\ l_i \in \mathcal{L} \text{ and } n \in \mathbb{N} \right\}$$

is dense in A, too.

The generalizations towards range Banach spaces Y is actually simple, also the analogous result for Lip^k.

vector, norm, complete(converge)

Signature on $\mathrm{Lip}([0,1];\mathbb{R}^d)$

Lipschitz curves (starting at 0) are a dual space of a Banach space (see work of Nigel Kalton and Sten Kaijser) and carry therefore a weak-$*$-topology, which constitutes a weighted space (E,ρ) where we take $\rho(u) = \exp(\|u\|^2_{\mathrm{Lip}})$.

For every Lipschitz curve u starting at 0 we can define signature of the curve extended by time, whose span provides us with a point separating subalgebra A of $B^\rho(E)$ satisfying the condition that $\exp(|l(u)|) \in B^\rho(E)$ for all $l \in E'$. Whence A is dense in $B^\rho(E)$. (analogous for rough path spaces and their corresponding weak-$*$-topologies (which differs from rough path norms!).

Results of Chevyrev-Oberhauser

- Robin Giles' strict topology is used on $C_b(E)$, where E is a topological space. This (locally convex) topology is weaker than the uniform topology but stronger then convergence on compacts in E. In particular the dual space is the space of finite Borel measures on E.
- The use of Stone-Weierstrass theorems on $C_b(E)$ demands for tensor normalization, which in turn interferes with the algebraic properties.
- This is the starting point for kernelizations, which are an extremely useful tool for analyzing and calculating approximations by signature basis.

Results of Chevyrev-Oberhauser

- Robin Giles' strict topology is used on $C_b(E)$, where E is a topological space. This (locally convex) topology is weaker than the uniform topology but stronger then convergence on compacts in E. In particular the dual space is the space of finite Borel measures on E.
- The use of Stone-Weierstrass theorems on $C_b(E)$ demands for tensor normalization, which in turn interferes with the algebraic properties.
- This is the starting point for kernelizations, which are an extremely useful tool for analyzing and calculating approximations by signature basis.

Results of Chevyrev-Oberhauser

- Robin Giles' strict topology is used on $C_b(E)$, where E is a topological space. This (locally convex) topology is weaker than the uniform topology but stronger then convergence on compacts in E. In particular the dual space is the space of finite Borel measures on E.
- The use of Stone-Weierstrass theorems on $C_b(E)$ demands for tensor normalization, which in turn interferes with the algebraic properties.
- This is the starting point for kernelizations, which are an extremely useful tool for analyzing and calculating approximations by signature basis.

New results

- tensor normalizations are unnecessary when working in $B^\rho(E)$. Still the dual space is a well understood space of Borel measures (those integrating ρ) and Stone-Weierstrass works in the particular case of the algebra generated by signatures.
- randomized signature can be considered a path space counterpart of a randomly initialized shallow networks. This opens the door for random feature analysis in the sense of Nicholas Nelsen and Andrew Stuart.
- this complements the pathwise Johnson-Lindenstrass inspired proof where randomized signature can be useful.

New results

- tensor normalizations are unnecessary when working in $B^\rho(E)$. Still the dual space is a well understood space of Borel measures (those integrating ρ) and Stone-Weierstrass works in the particular case of the algebra generated by signatures.
- randomized signature can be considered a path space counterpart of a randomly initialized shallow networks. This opens the door for random feature analysis in the sense of Nicholas Nelsen and Andrew Stuart.
- this complements the pathwise Johnson-Lindenstrass inspired proof where randomized signature can be useful.

New results

- tensor normalizations are unnecessary when working in $B^{\rho}(E)$. Still the dual space is a well understood space of Borel measures (those integrating ρ) and Stone-Weierstrass works in the particular case of the algebra generated by signatures.
- randomized signature can be considered a path space counterpart of a randomly initialized shallow networks. This opens the door for random feature analysis in the sense of Nicholas Nelsen and Andrew Stuart.
- this complements the pathwise Johnson-Lindenstrass inspired proof where randomized signature can be useful.