Mathematics for New Technologies in Finance

Solution sheet 2

Exercise 2.1 (Weierstrass theorem [1])

(a) Construct a sequence of polynomials converges pointwisely but not uniformly on $[0, 1]$.

(b) Construct a sequence of polynomials converges uniformly to $x \mapsto |x|$ on $[-1, 1]$. (Hint: Corollary 2.3. in [1])

(c) Prove that ReLU can be approximated uniformly by polynomials on $[-1, 1]$.

(d) Use the universal approximation theory of shallow neural networks on $[0, 1]$ to prove the Weierstrass theorem.

Solution 2.1

(a) Consider the functions $f_n(x) = x^n$ for $x \in [0, 1]$, or the functions $f_n = \mathbb{I}_{(0, \frac{1}{2^n})}$.

(b) Consider the following map

$$p_{n+1}(x) = p_n(x) + \frac{1}{2}(x - p_n^2(x)), \tag{1}$$

which is a contraction on $[0, 1)$ and the special case $x = 1$ is obvious. For more details, we refer to Corollary 2.3. in [1].

(c) $g(x) = \frac{1}{2}(x + |x|)$

(d) Since ReLU can be approximated uniformly by polynomials on $[0, 1]$, composition of affine function and ReLU can be uniformly by polynomials on $[0, 1]$. Thus, shallow neural networks can be uniformly by polynomials on $[0, 1]$. Therefore, by UAT, polynomials can uniformly approximate any continuous function on $[0, 1]$.

Exercise 2.2 (Networks on discrete path spaces)

(a) Describe the space of paths $\omega : \{1, \ldots, T\} \to \mathbb{R}^d$ as $\mathbb{R}^{dT}$.

(b) Describe a shallow neural network, which depends on value at time t and on path information up to time t. Formulate a universal approximation theorem in this setting.

Solution 2.2

(a) Maps from $\{1, \ldots, T\}$ to $\mathbb{R}^d$ expressed by $\mathbb{R}^{dT}$.

(b) A neural network with input space $\mathbb{R}^{dt}$ for fixed t, a neural network with input space at least $\mathbb{R}^{dT}$ (might be larger if allow duplicated information in input space). UAT for path space is concerning universal approximation of continuous functional on path spaces e.g. the running max of a discrete path.

Exercise 2.3 (Linear Operators) Let K be a compact subset of $\mathbb{R}^d$.

(a) Let μ be a finite Borel measure on K. Prove that

$$\mathcal{L}_\mu(f) := \int_K f(x)\mu(dx) \tag{2}$$

for $f \in C(K, \mathbb{R})$ is a bounded linear functional.

(b) Let $\mathcal{L}$ be a positive linear functional on $C(K, \mathbb{R})$, i.e. $\mathcal{L}(f) \geq 0$ for $f \geq 0$. Prove that $\mathcal{L}$ is continuous.

Solution 2.3

(a) $\mathcal{L}_\mu$ is linear by the linearity of the integral. We need to show that $\mathcal{L}_\mu$ is bounded. f is bounded, as f is continuous on K and K is compact. In addition, as $\mu(K) < \infty$, there exists $C \in \mathbb{R}$ such that $\mu(K) = C$. Hence

$$\mathcal{L}_\mu(f) = \int_K f(x)\mu(dx) \leq \int_K ||f||_\infty \mu(dx) \leq ||f||_\infty \mu(K) = ||f||_\infty C.$$

We have shown that there exists $C \in \mathbb{R}$ such that

$$\mathcal{L}(f) \leq ||f(x)||_\infty C, \forall f \in C(K, \mathbb{R}).$$

So $\mathcal{L}$ is bounded.

(b) We start by giving a reminder of the Riesz-Markov-Kakutani representation theorem.

Theorem 1 *Riesz-Markov-Kakutani representation theorem* Let X be a locally compact Hausdorff space, and $\mathcal{L}$ a positive linear functional on $C_c(X)$. Then there exists a unique positive Borel measure μ on X such that

$$\mathcal{L} = \int_X f(x)\mu(dx)$$

for every $f \in C_c(X)$, and which has the following properties for some M containing the Borel $\delta-$algebra on X:

1. $\mu(K) < \infty$ for every compact set $K \subset X$
2. For every $E \in M$, we have $\mu(E) = \inf\{\mu(V) : E \subset V, V \text{open}\}$
3. The relation $\mu(E) = \sup\{\mu(K) : K \subset E, K \text{compact}\}$ holds for every open set E, and for every $E \in M$ with $\mu(E) < \infty$
4. If $E \in M$, $A \subset E$, and $\mu(E) = 0$, then $A \in M$.

As $\mathcal{L}$ is positive linear functional, by Riesz-Markov-Kakutani representation theorem, there exists a unique measure μ such that the functional $\mathcal{L}$ on f is defined as $\mathcal{L}(f) := \int_K f(x)\mu(dx)$. Let a sequence of functions f_n in $C(K, \mathbb{R})$ converges uniformly to a function $f \in C(K, \mathbb{R})$, we have for any $\epsilon > 0$, there exists a positive integer N such that for all $n \geq N$ and $x \in K$, $|f_n(x) - f(x)| < \epsilon$. Since K is compact and f is continuous, f is also bounded on K, i.e., there exists a constant M such that $|f(x)| \leq M$ for all $x \in K$. Consequently, for all $n \geq N$, $|f_n(x)| \leq |f_n(x) - f(x)| + |f(x)| < \epsilon + M$. This implies $|f_n(x)|$ is bounded by $\epsilon + M$ for all $n \geq N$ and $x \in K$. Let $g_n(x) = max(|f_n(x)|, |f(x)|)$, we can see g_n is a bounded continuous function on compact set K, hence g_n is integrable. Thus we can apply dominated convergence theorem: If $f_n(x) \geq 0, f_n(x)$ converges to $f(x)$ pointwisely for all $x \in K$, and $|f_n(x)| \leq g_n(x)$ for all n and x, where $g_n(x)$ is integrable, then

$$\lim_{n\to\infty} \int_K f_n(x)\mu(dx) = \int_K f(x)\mu(dx)$$

So we have

$$\lim_{n\to\infty} \mathcal{L}(f_n) = \mathcal{L}(f)$$

It proves $\mathcal{L}$ is continuous.
Other proofs are also possible, for instance recalling that a linear operator between normed spaces is bounded if and only if it is continuous, and proceeding by contradiction.

Exercise 2.4 (Point-separating families)

(a) Let K be a compact subset of $\mathbb{R}^d$. Prove that

$$\mathcal{F} := \left\{ \mathcal{C}(K, \mathbb{R}) \ni f \mapsto \sum_{i=1}^{n} \lambda_i f(x_i) \mid \lambda_i \in \mathbb{R}, n \in \mathbb{N}, x_i \in K, i = 1, 2, ..., n \right\} \tag{3}$$

is point separating and additive.

(b) Prove that

$$\mathcal{F} := \left\{ \mathcal{C}_0^1([0,1], \mathbb{R}) \ni X \mapsto \sum_{i=1}^{n} \lambda_i \int t^i dX_t \in \mathbb{R} \colon \forall \lambda_i \in \mathbb{R}, n \in \mathbb{N} \right\}$$

is a point-separating vector space. $\mathcal{C}_0^1([0,1], \mathbb{R})$ is the space of the $\mathcal{C}^1$ function f on $[0,1]$ with $f(0) = 0$.

Solution 2.4

(a) Let f and g be distinct functions in $\mathcal{C}(K, \mathbb{R})$. Since they are distinct, there must exist at least one point $\bar{x}$ such that $f(\bar{x}) \neq g(\bar{x})$. Select now the element $F \in \mathcal{F}$ such that $n = 1$, $\lambda_1 = 1$ and $x_1 = \bar{x}$. Then, $F(f) \neq F(g)$. The additivity on $\mathcal{F}$ comes from

$$F(f+g) = \sum_{i=1}^{n} \lambda_i (f+g)(x_i) = \sum_{i=1}^{n} \lambda_i f(x_i) + \lambda_i g(x_i) = \sum_{i=1}^{n} \lambda_i f(x_i) + \sum_{i=1}^{n} \lambda_i g(x_i) = F(f) + F(g), \forall F \in \mathcal{F}.$$

(b) Vector space holds directly from the definition. So we remain to show point-separating. Let us consider $Z \in \mathcal{C}_0^1([0,1], \mathbb{R})$ s.t.

$$\int \sum_{i=1}^{n} \lambda_i t^i dZ_t = 0, \quad \forall \lambda_i \in \mathbb{R}, n \in \mathbb{N}.$$

An elementary approach is using universal approximation of polynomials. Since Z' is continuous on $[0,1]$, it can be universally approximated by polynomials, and therefore we have

$$\int_0^1 (Z_t')^2 dt = \int_0^1 Z_t dZ_t = \lim_{n\to\infty} \int \sum_{i=1}^{n} \lambda_i t^i dZ_t = 0. \tag{4}$$

This implies that $Z = 0$ because it starts from 0, which completes the proof.
It worth noticing that this essentially relies on the fact that Z' is continuous. But we can actually make the proof more general by considering function X which are only L-Lipschitz and starting from 0, and then a more general proof can be done by fourier analysis. Since $\sin(m\pi t)$ and $\cos(m\pi t)$ for all $m \in \mathbb{N}$ are uniformly approximated by polynomial on $[0,1]$. We have for all $m \in \mathbb{N}$

$$\int \sin(mt) dZ_t = \int \cos(mt) dZ_t = 0 \tag{5}$$

Then we define a sign measure $\mu(dt) = Z'_t dt$ (Because by Rademacher's Lipschitz function is almost everywhere differentiable and here we even know that $|Z'_t| \le L$ almost surely), then for all $m \in \mathbb{N}$

$$\int \sin(mt) d\mu = \int \cos(mt) d\mu = 0. \tag{6}$$

Then by fourier analysis we know $\mu = 0$ so Z is constant, which is actually 0 because $Z(0) = 0$. This proof uses the same idea used in the proof of universal approximation theory of neural network by G. Cybenko.

Exercise 2.5 (Controlled ODEs as features on the path space) We aim to demonstrate that controlled ODEs define (non-linear) features on a path space, which we shall fix to $\mathcal{C}^1([0,T], \mathbb{R}^d)$). See notebook 1 for details.

Solution 2.5 See solution notebook 1.

References

[1] Sameer Chavan. Problems and notes: uniform convergence and polynomial approximation.