Kolmogorovova veta a neurónové siete

Gabriela Andrejková

Ústav informatiky, PF UPJŠ, Košice

November 23, 2015

História

- výsledok ruského matematika A. N. Kolmogorova, 1957,
- ukazuje nový pohľad na možnosti dopredných neurónových sietí (NS), špeciálne na funkcie, ktoré sú nimi počítané,
- je základom pre aproximačné schopnosti NS,
- Kolmogorova veta obsahuje presné výrazy pre vyjadrenie n-dimenzionálnej spojitej funkcie pomocou 1-dimenzionálnych spojitých funkcií.

Opakovanie známych výsledkov

- Každý (separovateľný metrický) priestor X dimenzie menšej alebo rovnej n môže byť topologický vnorený do E^{2n+1} (Euclidovský $(2n+1)$-dimenzionálny priestor).
- To znamená, že trieda konečných dimenzionálnych priestorov je topologicky identická s triedou podmnožín Euclidovských priestorov.

Označenie

- $C(X)$ – priestor všetkých spojitých funkcií nad kompaktným priestorom X so supremovou normou:

$$\|f\| = sup_{x \in X} |f(x)|$$

.

- I^1 – interval $\langle 0, 1 \rangle$,
- I^n je karteziánsky súčin n I^1 intervalov, t.j. $I^n = \underbrace{\langle 0, 1 \rangle \times \cdots \times \langle 0, 1 \rangle}_{n\text{-krát}}$.

Doklady akademii nauk SSSR, 1957

Theorem (1. Kolmogorov)

Pre ľubovoľné $n, n \geq 2$*, existujú spojité reálne funkcie* $\psi^{p,q}$ *definované na* $I^1 = \langle 0, 1 \rangle$ *také, že ľubovoľnú spojitú reálnu funkciu* n*-premenených je možné vyjadriť v tvare:*

$$f(x_1, \ldots, x_n) = \sum_{q=1}^{2n+1} \chi^q \left[\sum_{p=1}^{n} \psi^{pq}(x_p) \right], \tag{1}$$

kde $\chi^q(y), q = 1, 2, \ldots, 2n+1$ *sú spojité funkcie jednej premennej.*

Základné myšlienky dôkazu - 4 kroky

1. Skoro celý I^n pokryjeme navzájom disjunktnými n-rozmernými kockami.
2. Potom zostrojíme spojité, monotónne rastúce funkcie ψ^{pq}, ktoré sú definované na I^1 tak, aby pre dané q, funkcia n-premenných $\sum\limits_{p=1}^{n} \psi^{pq}(x_p)$ zobrazila n-rozmerné kocky do navzájom disjunktných intervalov.
3. Na týchto intervaloch definujeme funkcie χ_r^q podľa predpisu a mimo intervalov dodefinujeme na spojité funkcie.

Základné myšlienky dôkazu - 4 kroky

4. Ukážeme, že funkcie χ^q, ktoré potrebujeme, sú funkcie $\chi^q = \lim_{r\to\infty} \chi^q_r$.
   Potom sa už ľahko dokáže, že funkcie χ^q a ψ^{pq} vyhovujú vzťahu (1).

Dôkaz je konštrukčný a veľa hovorí o vlastnostiach funkcií ψ^{pq} a χ^q.

I. Konštrukcia funkcií ψ^{pq}, $1 \leq p \leq n$; $1 \leq q \leq 2n+1$.

Nech k, i sú indexy také, že: $k = 1, 2, \ldots$; $m_k = (9n)^k + 1$; $1 \leq i \leq m_k$.
Ku každej trojici indexov q, k, i skonštruujeme interval (Fig. 1.):

$$A^q_{k,i} = \left\langle \frac{1}{(9n)^k}\left(i - 1 - \frac{q}{3n}\right), \frac{1}{(9n)^k}\left(i - \frac{1}{3n} - \frac{q}{3n}\right)\right\rangle. \qquad (2)$$

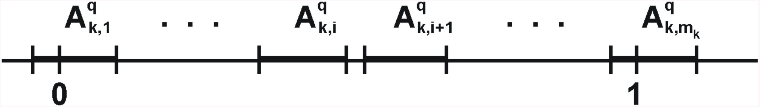


Fig. 1. Intervaly $A^q_{k,i}$ for $i = 1, \ldots, m_k$.

- Dĺžka intervalu $A_{k,i}^q$ je $\frac{1}{(9n)^k}(1-\frac{1}{3n})$. Je možné použiť aj iné metódy na rozdelenie intervalu I^1 na podintervaly.
- Pre ľubovoľné pevne zvolené hodnoty k a q posunutím intervalu $A_{k,i}^q$ o vzdialenosť $\frac{1}{(9n)^k}$ dostaneme interval $A_{k,i+1}^q$.
- Intervaly $A_{k,i}^q$ a $A_{k,i+1}^q$ sú disjuntné, vzdialenosť medzi nimi (dĺžka medzery) je $\frac{1}{3n(9n)^k}$.
- Pre každé i, i', $i \neq i'$, je $A_{k,i}^q \cap A_{k,i'}^q = \emptyset$. Intervaly $A_{k,i}^q$, kde q, k sú pevné, $i = 1, \ldots, m_k$, pokrýva takmer celý interval I^1, okrem medzier medzi nimi.

Konštrukcia funkcií ψ^{pq}.

Nech $S^q_{k,i_1\dots i_n}$ je karteziánsky súčin intervalov A^q_{k,i_p}, pre $p = 1, \dots, n$,

$$S^q_{k,i_1\dots i_n} = A^q_{k,i_1} \times \dots \times A^q_{k,i_n},$$

Je zrejmé, že pre rôzne n-tice, $(i_1, \dots, i_n)$, $(j_1, \dots, j_n)$ platí

- $S^q_{k,i_1\dots i_n} \cap S^q_{k,j_1\dots j_n} = \emptyset$ a
- $S^q_k = \bigcup\limits_{(i_1,\dots,i_n)} S^q_{k,i_1\dots i_n}$ je *zjednotenie n-dimenzionalných kociek*, ktoré pokrýva takmer celý I^n.

Theorem (2.)

Je možné nájsť konštanty $\lambda_{k,i}^{pq}$ a ε_k také že spĺňajú nasledujúce podmienky:

1. $\lambda_{k,i}^{pq} < \lambda_{k,i+1}^{pq} \leq \lambda_{k,i}^{pq} + \frac{1}{2^k}$
2. $\lambda_{k,i}^{pq} \leq \lambda_{k+1,i'}^{pq} \leq \lambda_{k,i}^{pq} + \varepsilon_k - \varepsilon_{k+1}$, *if* $A_{k,i}^{q} \cap A_{k+1,i'}^{q} \neq \emptyset$
3. *Intervaly* $\Delta_{k,i_1\ldots i_n}^{q} = \left\langle \sum_{p=1}^{n} \lambda_{k,i_p}^{pq}, \sum_{p=1}^{n} \lambda_{k,i_p}^{pq} + n.\varepsilon_k \right\rangle$ *pri pevnom* p *a* q*, sú navzájom disjunktné.*

Z (2) ľahko vidieť, že $\varepsilon_k \leq \frac{1}{2^k}$.

Nech $\Delta^q_{k,i_1\dots i_n}$ a $\Delta^q_{k,i_1\dots i_{n-1}i'_n}$, sú dva intervaly kde $i'_n = i_n + 1$. Podľa 3. a 1. bodu v leme 2, tieto intervaly môžu byť zapísané v tvare:

$$\Delta^q_{k,i_1\dots i_n} = \left\langle \sum_{p=1}^{n} \lambda^{pq}_{k,i_p}, \sum_{p=1}^{n} \lambda^{pq}_{k,i_p} + n.\varepsilon_k \right\rangle$$

$$\Delta^q_{k,i_1\dots i_{n-1}i'_n} = \left\langle \sum_{p=1}^{n} \lambda^{pq}_{k,i_p} + d, \sum_{p=1}^{n} \lambda^{pq}_{k,i_p} + d + n.\varepsilon_k \right\rangle,$$

kde $d = \lambda^{pq}_{k,i_n+1} - \lambda^{pq}_{k,i_n}$.

Podľa 3. bodu v (2), tieto intervaly sú disjunktné a podľa 1. bodu platí, že: $0 < d \leq \frac{1}{2^k}$. Preto musí byť splnené tiež

$$\sum_{p=1}^{n} \lambda_{k,i_p}^{pq} + n.\varepsilon_k < \sum_{p=1}^{n} \lambda_{k,i_p}^{pq} + d$$

a $\varepsilon_k < \frac{d}{n}$.

Nech p a q sú dva indexy. Čo je možné povedať o funkciách ψ^{pq}?
Keďže platia nerovnosti $\varepsilon_k < \frac{d}{n}$ a $d \leq \frac{1}{2^k}$, máme:

4. $\varepsilon_k \leq \frac{1}{2^k}$.

Na základe vyššie uvedených vlastností intervalov $A_{k,i}^{pq}$ a konštant $\lambda_{k,i}^{pq}$, ε_k je možné dokázať nasledujúcu vetu.

Theorem (3.)

Pri pevne daných p a q podmienka

5. $\lambda_{k,i}^{pq} \leq \psi^{pq}(x) \leq \lambda_{k,i}^{pq} + \varepsilon_k$ *pre* $x \in A_{k,i}^{q}$

jednoznačne definuje spojitú funkciu $\psi^{pq}(x)$ na intervale I^1. ⊙

Pretože platí $\lambda_{k,i}^{pq} < \lambda_{k,i+1}^{pq}$, funkcia ψ^{pq} bude na intervale $A_{k,i+1}^{q}$ nadobúdať väčšie hodnoty než na intervale $A_{k,i}^{q}$ (Fig. 2).

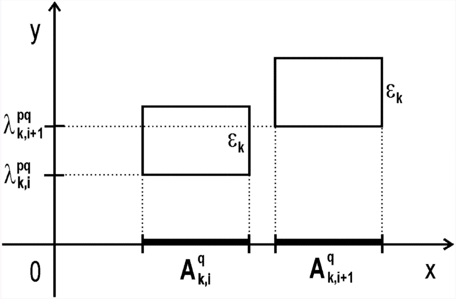


Fig. 2. Intervaly $A_{k,i}^{q}$ *a* $A_{k,i+1}^{q}$ *a konštanty* $\lambda_{k,i}^{pq}$ *a* $\lambda_{k,i+1}^{pq}$.

- Zvýšením hodnoty indexu k o 1, je skonštruovaný jemnejší rozklad intervalu I^1 na menšie podintervaly.
- Z 2. bodu v (2) je známe, že $\lambda_{k,i}^{pq} \leq \lambda_{k+1,i'}^{pq} \leq \lambda_{k,i}^{pq} + \varepsilon_k - \varepsilon_{k+1}$, pre
$$A_{k,i}^q \cap A_{k+1,i'}^q \neq \emptyset.$$
- Pravá nerovnosť implikuje, že
$$\lambda_{k+1,i'}^{pq} + \varepsilon_{k+1} \leq \lambda_{k,i}^{pq} + \varepsilon_k.$$
- Z 1. bodu a z platnosti $\varepsilon_k > 0$ vyplýva
$$\lambda_{k,i}^{pq} \leq \lambda_{k+1,i'}^{pq} < \lambda_{k+1,i'}^{pq} + \varepsilon_{k+1}.$$

- A pre každú dvojicu intervalov $A^q_{k,i}$ a $A^q_{k+1,i'}$, ktoré majú spoločné body platí

$$\lambda^{pq}_{k,i} \leq \lambda^{pq}_{k+1,i'} < \lambda^{pq}_{k+1,i'} + \varepsilon_{k+1} \leq \lambda^{pq}_{k,i} + \varepsilon_k.$$

- Hodnoty funkcií ψ^{pq} na intervale $A^q_{k,i}$ môžu byť vypočítané presnejšie na intervaloch $A^q_{k+1,i'}$, ktoré majú nejaké spoločné body s intervalom $A^q_{k,i}$ (Fig. 3).
- Funkcie ψ^{pq} môžu byť definované ako po častiach konštantné funkcie na malých intervaloch (znamená to ako schodovité funkcie).
- Je nutné ukázať, že funkcia ψ^{pq} je spojitá v každom bode intervalu I^1.

Z 3. a 5. bodu v (2) a (3) vyplýva:

6. $\sum_{p1}^{n} \psi^{pq}(x_p) \in \Delta^q_{k,i_1\dots i_n}$ for $(x_1, \dots x_n) \in S^q_{k,i_1\dots i_n}$.

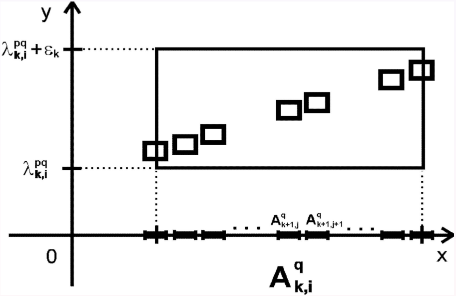


Fig. 3. Hodnoty funkcie ψ^{pq} na intervaloch $A^q_{k,i}$ a $A^q_{k+1,i'}$.

Funkcie χ^q budú konštruované pomocou funkcií χ^q_r ako limitné funkcie $\chi^q = \lim_{r\to\infty} \chi^q_r$. Nech:

$$f_r(x_1, \ldots, x_n) = \sum_{q=1}^{2n+1} \chi^q_r \left(\sum_{p=1}^{n} \psi^{pq}(x_p) \right) \tag{3}$$

$$M_r = \sup_{I^n} |f - f_r|. \tag{4}$$

Je možné ukázať, že pre $r \to \infty$ funkcie f_r sa blížia ku funkcii f, a to tak, že $\lim_{r\to\infty} f_r = f$.

Pre $r = 0$, definujeme $\chi^q_0(x) = 0$.

Z definície χ_0^q vyplýva:

$$f_0 = 0, \quad M_0 = \sup_{I^n} |f|.$$

Funkcie χ_r^q pre $r > 0$ budú definované rekurentne.

Máme definované spojité funkcie χ_{r-1}^q, f_{r-1} a máme dané k_{r-1} (k_{r-1} určuje rozklad I^n na n-dimenzionálne kocky).

Kedže dĺžky strán kociek $S^q_{k,i_1\dots i_n}$ pre $k \to \infty$ sa blížia k nule, vieme vybrať k_r také veľké, aby na ľubovoľnej n-dimenzionálnej kocke $S^q_{k_r,i_1\dots i_n}$ pre všetky $\bar{x}, \bar{y} \in S^q_{k_r,i_1\dots i_n}$ bola splnená nasledujúca nerovnosť:

$$|(f - f_{r-1})(\bar{x}) - (f - f_{r-1})(\bar{y})| \leq \frac{1}{2n+2} M_{r-1}.$$

Nech $\xi^q_{k,i}$ je ľubovoľný bod intervalu $A^q_{k,i}$.
$(\xi^q_{k,i_1}, \ldots, \xi^q_{k,i_n})$ je bod n-dimenzionálnej kocky $S^q_{k,i_1\ldots i_n}$.

Funkcia $\chi^q_r(y)$ bude definovaná na intervale $\Delta^q_{k,i_1\ldots i_n}$ nasledovne:

$$\chi^q_r(y) = \chi^q_{r-1}(y) + \frac{1}{n+1}\left(f(\xi^q_{k,i_1}, \ldots, \xi^q_{k,i_n}) - f_{r-1}(\xi^q_{k,i_1}, \ldots, \xi^q_{k,i_n})\right). \tag{5}$$

Z (5) dostávame:

$$|\chi^q_r(y) - \chi^q_{r-1}(y)| = \frac{1}{n+1}|f(\xi^q_{k,i_1}, \ldots, \xi^q_{k,i_n}) - f_{r-1}(\xi^q_{k,i_1}, \ldots, \xi^q_{k,i_n})|$$

$$\leq \frac{1}{n+1}\sup_{I^n}|f - f_{r-1}| = \frac{1}{n+1}M_{r-1}.$$

a nasledujúca nerovnosť (6) je splnená:

$$|\chi_r^q(y) - \chi_{r-1}^q(y)| \leq \frac{1}{n+1} M_{r-1}. \tag{6}$$

V medzerách medzi intervalmi $\Delta^q_{k,i_1,\ldots,i_n}$ funkcia $\chi_r^q(y)$ bude definovaná ako spojitá funkcia tak, aby bola splnená nerovnosť (6).
Rozdiel $f(\overline{x}) - f_r(\overline{x})$ môže byť ohraničený pomocou vzťahov (7) a (8) a vytvára klesajúcu postupnosť pre $r \to \infty$.

$$f(\overline{x}) - f_r(\overline{x}) = f(\overline{x}) - f_{r-1}(\overline{x}) - \sum_{q=1}^{2n+1} \left(\chi_r^q\left(\sum_{p=1}^{n} \psi^{pq}(x_p) \right) - \chi_r^q\left(\sum_{p=1}^{n} \psi^{pq}(x_p) \right) \right) \tag{7}$$

A tiež platí

$$|f(\bar{x}) - f_r(\bar{x})| \leq \frac{2n+1}{2n+2} M_{r-1}, \tag{8}$$

pre ľubovoľné $\bar{x} = (x_1, x_2, \ldots, x_n) \in I^n$.
Z (8) dostávame

$$M_r \leq \frac{2n+1}{2n+2} \cdot M_{r-1}$$

$$M_r \leq \left(\frac{2n+1}{2n+2}\right)^r \cdot M_0. \tag{9}$$

Pomocou (4) a (9) je možné dokázať, že $\lim\limits_{r\to\infty} f_r = f$ a postupnosť funkcií χ_r^q konverguje rovnomerne k funkcii χ^q.

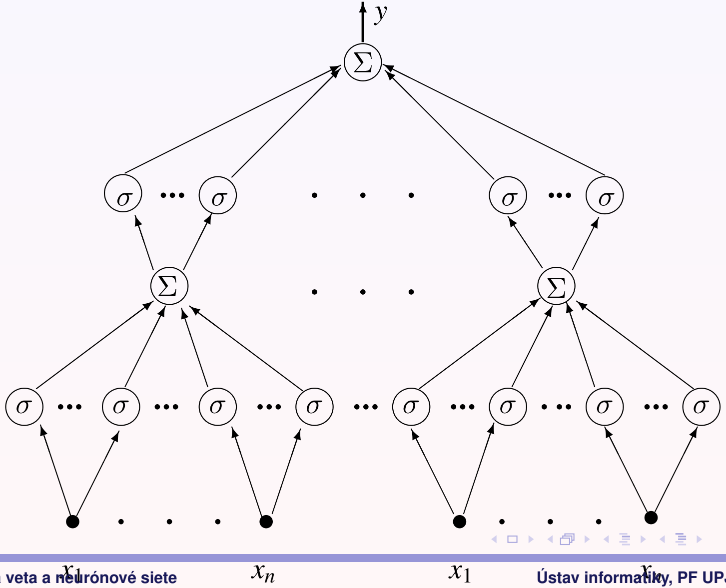
Rovnica (1) môže byť interpretovaná ako 3-vrstvová dopredná neurónová sieť, kde

- k-tý skrytý neurón počíta funkciu $z_q = \sum_{p=1}^{n} \psi^{pq}(x_p)$,
- výstupné neuróny počítajú funkcie $\sum_{q=1}^{2n+1} \chi^q(z_q)$ a
- z_k je výstup skrytej vrstvy.

Nie je tu však povedané, aký tvar majú mať funkcie χ^q, ψ^{pq}. Kolmogorovovu vetu vylepšil Sprecher, ktorý sformuloval iný tvar vnútornej sumy vo vzťahu (1).

Prvou prácou, v ktorej bola študovaná univerzálna aproximácia dopredných NS bola práca Hecht-Nielsena.
Kurková dokázala, že pri úprave (1) na aproximačnú nerovnosť, je potrebné riešiť nasledujúce:

1. Funkcie ψ^{pq} vysoko hladké.
2. Funkcie χ^q závisia od f a nie sú reprezentovateľné v parametrickom tvare.


y
Σ
σ
σ
Σ
Σ
σ
σ
x_1
x_n
x_1
x_n

Theorem (Kurková, 1991)

Nech $n \geq 2$ *je prirodzené číslo,* $\sigma : R \rightarrow [0,1]$ *je sigmoidálna funkcia,* $f : [0,1] \rightarrow R$ *je spojitá funkcia a* ε *je kladné reálne číslo. Potom existuje prirodzené číslo* k *a schodovité funkcie typu* σ $\psi^{p,i}, \phi^i, i = 1 \ldots, k, p = 1, \ldots, n$ *také,že pre každé* $(x_1, \ldots, x_n) \in [0,1]^n$ *platí*

$$\left| f(x_1, \ldots, x_n) - \Sigma_{i=1}^{k} \phi^i \left(\Sigma_{p=1}^{n} \psi^{p,i}(x_p) \right) \right| < \varepsilon. \quad \odot \tag{10}$$

Označenie

$$w(h, \delta) = sup\{h(\overline{x}) - h(\overline{y})\},$$
$$\overline{x} = (x_1, \ldots, x_n), \overline{y} = (y_1, \ldots, y_n) \in I^n, |x_i - y_i| < \delta, i = 1, \ldots, n,$$
$$\|f\|_L^\infty = sup_{x \in R}\{f(x)\} < \infty.$$

Theorem (Kurková, 1992)

Nech $n \in N, n \geq 2$, *nech* $\sigma : R \to I$ *je sigmoidálna funkcia,* $f \in C(I^n)$ *a nech* ε *je kladné reálne číslo. Potom pre každé* $m \in N$ *také, že* $m \geq 2n+1$ *a* $\frac{n}{m-n} + v < \frac{\varepsilon}{\|f\|}$ *a* $\omega_f(1/m) < v\frac{m-n}{2m-3n}$ *pre nejake kladné číslo* $v \in R$ *môže byť funkcia* f *aproximovaná s presnosťou* ε *pomocou doprednej neurónovej siete s dvoma skrytými vrstvami, pričom v prvej skrytej vrstve má sieť* $nm(m+1)$ *neurónov a v druhej skrytej vrstve má* $m^2(m+1)^n$ *neurónov s aktivačnou funkciou* σ*. Navyše všetky váhy a prahy siete s výnimkou váh spájajúcich druhú skrytú vrstvu a výstupný neurón , sú pevné pre všetky funkcie* g*, pre ktoré platí:* $\|g\| \leq \|f\|$ *a zároveň* $\omega_g(\delta) \leq \omega_f(\delta)$ *pre ľubovoľné* $\delta > 0$. $\odot$

Zaujímavé výsledky poslednej vety

1. horné hranice na počet neurónov v skrytej vrstve, a
2. len váhy druhej skrytej vrstvy závisia od funkcie f.

Kurkovej výsledky vedú k návrhu nových učiacich algoritmov. Kurková a Neruda's charakterizovali funktionalitu ekvivalentných Gaussian Radial Basis Function (RBF) sietí.

References

1. Kolmogorov, A. N. *On the representation of continuous functions of several variables by superposition of continuous functions of one variable and addition*. Doklady Akademii Nauk USSR, No. 5, Vol. 114, 1957, pp. 679–681.
2. Kurková, V. *Kolmogorova veta is relevant*. Neural Computation, No. 3, 1991, pp. 617–622.
3. Kurková, V. *Kolmogorov's theorem and multilayer network*. Neural Networks, 5(3), 1992, pp. 501–506.

References

4. Kurková, V., Hlaváčková, K. *Approximation of continuous functions by RBF and KBF networks.* In Proceedings of ESANN'94, Brussels, 1994, pp. 167-174.
5. Lorentz, G. G. *Approximation of functions.* New York, Holt, Reihardt and Winston, 1966.
6. Nees, M. *Approximative versions of Kolmogorov's superposition.* J. Comput. Appl. Math., 54, 1994, pp. 277-278.
7. Šíma, J., Neruda, R.: *Theoretical Questions of Neural Networks*, Matfyzpress, Praha, 1996.