Pokroky matematiky, fyziky a astronomie

Věra Pohlová

Učení neuronových sítí jako inverzní úloha

Pokroky matematiky, fyziky a astronomie, Vol. 49 (2004), No. 3, 218–225

Persistent URL: http://dml.cz/dmlcz/141231

Učení neuronových sítí jako inverzní úloha

Věra Kůrková, Praha

1. Úvod

Matematické metody rozvinuté v průběhu několika století pro řešení fyzikálních inverzních problémů lze využít i v matematické teorii neuronových sítí. Pomocí nich můžeme popsat optimální řešení úlohy učení na základě příkladů a modelovat schopnost generalizace.

Na rozdíl od klasické umělé inteligence, která řeší kognitivní úlohy pomocí manipulací se symboly založených na pravidlech, *konekcionismus* nahrazuje namáhavé hledání pravidel *učením na základě příkladů*. Konekcionismus byl inspirován zjednodušenými představami o mozku tvořeném sítěmi vzájemně propojených neuronů (odtud název konekcionismus) s měnícími se propustnostmi jednotlivých synapsí. Výpočetním modelům, které konekcionismus používá, se říká *umělé neuronové sítě*. Tyto sítě mohou řešit úlohy, které lze reprezentovat jako transformace mezi různými druhy kódů. Např. systém NETtalk [19] „čte nahlas“ anglický text tak, že transformuje grafický kód na fonetický. NETtalk byl vytvořen jako alternativa ke komerčnímu expertnímu systému DECtalk, založenému na pravidlech výslovnosti a seznamu výjimek. Vzhledem k tomu, že anglická výslovnost závisí na kontextu, NETtalk potřebuje pro nalezení správného fonému pro výslovnost jednoho znaku anglického textu znát řetězec 7 znaků obsahující 3 předcházející a 3 následující znaky (například znak „o“ ve slovech „tough“ a „through“ odpovídá jiným fonémům). Znaky zpracovávaného textu jsou kódovány jako binární vektory délky 29 odpovídající 26 písmenům anglické abecedy a 3 typům mezer a hranic slov. NETtalk zobrazuje grafický kód tvořený binárními vektory délky 7×29 na fonetický kód, který řídí zvukový syntetizátor.

Flexibilita umělých neuronových sítí spočívá v množství proměnných *parametrů* (někdy se jim říká váhy), jejichž nastavení určuje transformaci vstupních dat na výstupní data (funkci vstup–výstup). Cílem učení je nastavit parametry sítě tak, aby síť v rámci určité tolerance přesnosti prováděla požadovanou transformaci. Vzhledem k tomu, že analytický popis funkce vstup–výstup není znám, parametry systému lze hledat pouze na základě vzorku správně zpracovaných dvojic vstup–výstup (tzv. trénovací množina). Učení je tedy hledání takové parametrizace sítě, která určuje transformaci vstup-výstup dobře aproximující empirická data z trénovací množiny.

RNDr. Věra Kůrková, DrSc. (1948), Ústav informatiky Akademie věd ČR, Pod Vodárenskou věží 2, 182 07 Praha 8, e-mail: vera@cs.cas.cz, http://www.cs.cas.cz/~vera

Předneseno na semináři Kognice a umělý život, Opava, 26. – 30. 5. 2004.

Tato práce vznikla s částečnou podporou grantu GA ČR 201/02/0428.

Podobnou úlohou požadující nalezení funkce odpovídající naměřeným astronomickým datům se zabývali počátkem 19. století Gauss a Legendre. Vyřešili ji pomocí *metody nejmenších čtverců*, která hledá funkci s minimálním součtem čtverců chyb na daných empirických datech. Legendre napsal v r. 1805 v Nouvelles méthodes pour la determination des orbit des cometes: *„Myslím, že ze všech principů, které mohou být navrženy, žádný není obecnější, exaktnější a snadněji použitelný než ten, který spočívá v minimalizaci součtu čtverců chyb.“* Gauss navíc ukázal, že metoda nejmenších čtverců má pozoruhodné statistické vlastnosti [4].

Úloha nalézt funkci aproximující empirická data patří do širší třídy tzv. *inverzních problémů* zabývajících se hledáním *neznámých příčin* (např. sil, tvarů funkcí či distribucí) *známých následků* (naměřených dat).

Závislost následků na příčinách je zpravidla popsána ve tvaru operátoru, v nejjednodušším případě lineárního, přiřazujícího příčinám (nazývaným *řešení*) následky (nazývané *data*). Řešení inverzních problémů je základním úkolem mnoha oblastí aplikovaného výzkumu, jako je např. lékařská diagnostika (tomografie), seismologie a meteorologie.

Hadamard v r. 1902 definoval pojem *korektního inverzního problému* (angl. well-posed) jako problému, který má pro *všechna data z daného prostoru jediné řešení závisející spojitě na datech*. Hadamard ukázal, že některé klasické fyzikální inverzní problémy jsou korektní. Domníval se, že nekorektní problémy (tj. problémy, které nemají pro všechna data řešení nebo řešení není jednoznačné anebo nezávisí spojitě na datech) nemají fyzikální smysl. Fyzika, kterou se zabýval Hadamard, byla fyzika 19. století, směřující k Laplaceovu ideálu jednoznačného, stabilního a deterministického popisu světa. V 19. století byly nekorektní problémy považovány za anomálie podobně jako spojité funkce bez derivace, Peanova křivka, Cantorovo diskontinuum a ďábelské schodiště, které vedly ve druhé polovině 20. století k vytvoření fraktální geometrie.

Podobný vývoj od anomálie k centru zájmu aplikované matematiky proběhl také v oblasti inverzních úloh. Postupně byla vyvinuta metodologie řešení nekorektních úloh založená na *Mooreově-Penroseově pseudoinverzi* a různých typech *regularizace*. Pro konečněrozměrné problémy, které nemají pro všechna data řešení, navrhl v r. 1920 Moore [13] metodu zobecněné inverze založenou na hledání tzv. *pseudořešení*. Jeho práce publikovaná pouze jako abstrakt zůstala nedoceněna a metoda pseudoinverze nebyla dále rozvíjena ani využívána až do jejího znovuobjevení Rogerem Penrosem v r. 1955 [16]. V 70. letech byla Mooreova-Penroseova metoda rozšířena i na inverzní problémy definované pomocí spojitých operátorů na nekonečněrozměrných Hilbertových prostorech [8].

Vzhledem k tomu, že empirická data podléhají chybám měření, podstatnou vlastností operátoru popisujícího inverzní problém je stabilita vzhledem k šumu. V mnoha inverzních úlohách malá změna dat vede k velké změně řešení nebo pseudořešení. Takovým úlohám se říká *špatně podmíněné*. Metoda *regularizace* vyvinutá v 60. letech 20. století je založena na myšlence, že se lze zbavit nejednoznačnosti a zlepšit stabilitu řešení, pokud kromě empirických dat vezmeme také v úvahu *konceptuální* *data* popisující nějakou globální vlastnost hledané funkce. To znamená, že omezíme prostor, kde hledáme řešení vyhovující naměřeným datům, pouze na řešení mající fyzikální smysl, tj. splňující nějakou předem danou podmínku, jako např. omezený výskyt oscilací vyšších frekvencí. Metodu objevilo nezávisle na sobě několik autorů, jednotný rámec jí dali Tichonov a Arsenin [20].

Rozvinutí metod pseudoinverze a regularizace a algoritmů na nich založených vedlo k vyřešení řady nekorektních úloh. Jednou z nejúspěšnějších aplikací řešení nekorektní úlohy je počítačová tomografie založená na výpočtu inverzní Radonovy transformace.[1]) Za její vynález byla v r. 1979 udělena Nobelova cena za medicínu. Počítačový obraz plošného řezu části těla je řešením inverzní úlohy s daty získanými rentgenováním podél velkého počtu přímek ze zobrazované plochy. Matematicky jde o rekonstrukci funkce dvou proměnných z integrálů této funkce podél přímek.

Počátkem devadesátých let 20. století byla regularizační metoda použita také v oblasti učení neuronových sítí [17]. Ukázalo se, že konceptuální data omezující prostor hledaných řešení jsou velmi vhodná pro modelování schopnosti *generalizace*, tj. schopnosti naučené sítě uspokojivě zpracovat i mnohá vstupní data, která nebyla součástí trénovací množiny.

Pro modelování konceptuálních dat, která jsou vyjádřena jako podmínky omezující oscilace (vysokofrekvenční filtry různých typů), jsou vhodným nástrojem *normy na speciálních Hilbertových prostorech*. Tyto prostory, nazývané Hilbertovy prostory s reprodukčním jádrem, obsahují pouze funkce splňující určité podmínky na oscilace, jejichž typy jsou určeny tzv. jádry. I když byla jádra intenzivně studována již v 19. století v souvislosti s integrálními rovnicemi, Hilbertovy prostory určené jádry definoval až v r. 1950 Aronszajn [1]. V 60. letech byly prostory s reprodukčním jádrem aplikovány ve statistice [15], později byly využity pro interpolaci dat pomocí splinů [21] a koncem 90. let též k modelování generalizace. Girosi [7] ukázal, že vysokofrekvenční filtry lze studovat v rámci teorie těchto prostorů a Cucker a Smale [5] je využili jako prostory hypotéz v teorii učení.

V tomto článku ukážeme, jak metody řešení matematických problémů z oblasti učení neuronových sítí zapadají do dlouhodobého vývoje matematických idejí původně motivovaných fyzikálními úlohami. Pomocí vhodně definovaného lineárního operátoru formulujeme úlohu učení jako inverzní problém. Je pozoruhodné, že Hilbertovy prostory určené jádrem, jejichž normy se tak dobře hodí pro vyjádření velikosti různých typů oscilací, jsou jedinou třídou nekonečněrozměrných Hilbertových prostorů, na níž operátor modelující učení splňuje podmínky potřebné pro aplikaci Mooreovy-Penroseovy pseudoinverze. Využití těchto prostorů tedy umožňuje elegantní popis řešení úlohy učení dané vzorkem empirických dat a současně nabízí možnost zlepšit stabilitu tohoto řešení přidáním konceptuálních dat modelovaných pomocí vhodného jádra. Zde uvádíme jen hlavní výsledky, podrobnější výklad je v [10] a [11].

---

[1]) *Poznámka redakce:* O Johannu Radonovi viz článek Fuchs, E., Netuka, I.: *Johann Radon (k stému výročí narození).* PMFA *33* (1988), 241–248.

2. Inverzní problémy

Pojem inverzní úlohy lze definovat v termínech funkcionální analýzy: Inverzní problém určený operátorem $A\colon (X, \|\cdot\|_X) \to (Y, \|\cdot\|_Y)$ mezi Banachovými prostory je úloha pro dané $g \in Y$ najít $f \in X$ takové, že $A(f) = g$. Prvky prostoru X se nazývají *řešení* a prvky prostoru Y *data*. Je-li Y konečněrozměrný, inverzní problém se nazývá problém s *diskrétními daty*.

Pokud pro každé $g \in Y$ existuje jediné řešení $f \in X$, inverzní problém se nazývá *korektní*. Pro korektní problém tedy existuje jediný inverzní operátor $A^{-1}\colon Y \to X$.

Je-li A spojitý, pak podle Banachovy věty (viz např. [6, s. 141]) je A^{-1} také spojitý. Ale ani spojitá závislost řešení na datech nezaručí vždy stabilitu vzhledem k šumu. Jako míra stability řešení inverzního problému se používá tzv. *číslo podmíněnosti* definované pro korektní problém jako $\operatorname{cond}(A) = \|A\|\,\|A^{-1}\|$. Dá se snadno ukázat, že je-li $A(f) = g$ a $A(f') = g'$, potom

$$\frac{\|f - f'\|_X}{\|f\|_X} \leqq \operatorname{cond}(A)\,\frac{\|g - g'\|_Y}{\|g\|_Y}\,.$$

Relativní odchylka řešení je tedy menší nebo rovna součinu čísla podmíněnosti a relativní odchylky dat. Číslo podmíněnosti je vždy větší nebo rovno 1. Pokud je malé, problém se nazývá *dobře podmíněný*, pokud je velké, hovoří se o *špatně podmíněném problému*.

Mnohé inverzní problémy jsou nekorektní: pro některá data nemají žádné řešení, nebo naopak mají mnoho různých řešení. I když řešení neexistuje, lze hledat alespoň jeho nejlepší aproximaci f^0, nazývanou *pseudořešení*, splňující rovnici

$$\|A(f^0) - g\|_Y = \min_{f \in X} \|A(f) - g\|_Y,$$

a v případě, že je těchto pseudořešení více, hledat mezi nimi *normální pseudořešení* f^+, které má nejmenší normu. Pokud pro všechna $g \in Y$ existuje normální pseudořešení f^+, pak může být definován *pseudoinverzní operátor* $A^+\colon Y \to X$ jako $A^+(g) = f^+$. Podobně jako pro korektní problémy lze též definovat *číslo podmíněnosti* jako $\operatorname{cond}(A) = \|A\|\,\|A^+\|$.

Jsou-li X a Y konečněrozměrné, lze operátor A popsat maticí a pseudoinverzní operátor odpovídá Mooreově-Penroseově pseudoinverzi této matice. Podobné vlastnosti má i pseudoinverze spojitých operátorů s uzavřenými obrazy, které jsou definované na nekonečněrozměrných prostorech [8], [3].

3. Minimalizace funkcionálu empirické chyby

Metodu nejmenších čtverců aplikovanou na učení na základě dat lze formulovat jako minimalizaci funkcionálu určeného vzorkem dat tvořeným dvojicemi vstup–výstup $z = \{(x_i, y_i) \in \Omega \times \mathbb{R}\colon i = 1, \ldots, m\}$ (m je velikost vzorku, Ω je neprázdná množina a $\mathbb{R}$ značí množinu reálných čísel). Tento funkcionál nazývaný *empirická chyba* je definován jako

$$\mathcal{E}_z(f) = \frac{1}{m} \sum_{i=1}^{m} \big(f(x_i) - y_i\big)^2.$$

Mnoho algoritmů učení neuronových sítí, jako např. zpětné šíření [22], minimalizuje hodnotu $\mathcal{E}_z\big(f(p)\big)$ na prostoru vektorů parametrů p sítí daného typu ($f(p)$ značí funkci vstup–výstup, kterou počítá síť určená vektorem parametrů p).

Minimalizaci tohoto funkcionálu lze formulovat jako inverzní úlohu určenou operátorem $L_x\colon X \to \mathbb{R}^m$ definovaným vztahem

$$L_x(f) = \left(\frac{f(x_1)}{\sqrt{m}}, \dots, \frac{f(x_m)}{\sqrt{m}}\right).$$

Protože $\mathcal{E}_z$ lze vyjádřit jako

$$\mathcal{E}_z = \left\| L_x - \frac{y}{\sqrt{m}} \right\|_2^2,$$

kde $\|\cdot\|_2$ značí l_2-normu na $\mathbb{R}^m$, problém minimalizace $\mathcal{E}_z$ na X je ekvivalentní úloze najít pseudořešení inverzního problému daného operátorem L_x pro data $y/\sqrt{m}$.

Zobecnění Mooreovy-Penroseovy metody pseudoinverze [3, str. 56–60], [8, str. 37–46] na operátory s nekonečněrozměrným definičním oborem umožňuje popis vlastností řešení pouze v případě, že operátor je *spojitý* a má *uzavřený obraz*. Pro aplikaci této teorie je tedy třeba najít Hilbertovy prostory funkcí, na nichž jsou evaluační funkcionály spojité. Obraz operátoru L_x definovaného na libovolném prostoru funkcí je uzavřený, protože je to podprostor konečněrozměrného prostoru $\mathbb{R}^m$.

Na rozdíl od mnoha inverzních problémů motivovaných fyzikálními ději není operátor L_x spojitý vzhledem k $\mathcal{L}_2$-normě. Evaluační operátory na podprostoru $\big(\mathcal{L}_2(\mathbb{R}^d), \|\cdot\|_2\big)$ obsahujícím pouze spojité funkce nemohou být spojité, protože nejsou omezené (operátor na normovaném prostoru je spojitý, právě když je omezený [6]). Například funkcionál evaluace v bodě 0 zobrazí omezenou posloupnost funkcí $\{n^d \mathrm{e}^{-(n\cdot\|x\|)^2}\}$ na neomezenou posloupnost reálných čísel.

Mooreovu-Penroseovu metodu pro nalezení minima funkcionálu empirické chyby tedy nemůžeme použít na podprostoru $\big(\mathcal{L}_2(\mathbb{R}^d), \|\cdot\|_2\big)$ tvořeném spojitými funkcemi (protože operátor L_x na něm není spojitý) ani na prostoru spojitých funkcí $\mathcal{C}(\mathbb{R}^d)$ se supremovou normou (protože to není Hilbertův prostor). Avšak zdaleka ne všechny funkce z těchto prostorů dávají rozumná řešení problému aproximace empirických dat použitelná i pro jiná data než data z trénovací množiny.

Naštěstí existují Hilbertovy prostory, které jsou jako šité na míru pro modelování učení na základě dat. Operátory evaluace jsou na nich spojité (to je jedna z jejich dvou ekvivalentních definic) a navíc jejich normy „měří“, jak rychle rostou vysokofrekvenční oscilace určitého typu. Pomocí podmínek omezujících velikost těchto norem můžeme tedy vyloučit ta řešení úlohy učení, která mají příliš velké oscilace.

Tyto Hilbertovy prostory se nazývají *prostory s reprodukčním jádrem* (anglicky reproducing kernel Hilbert spaces, proto se označují akronymem RKHS). Definoval je v r. 1950 Aronszajn jako Hilbertovy prostory bodově definovaných funkcí, na nichž jsou všechny evaluační funkcionály omezené. Aronszajn vycházel z práce mnoha generací matematiků, kteří studovali integrální operátory tvaru konvoluce s jádrem $L_K(f) = \int f(y)K(x,y)\,\mathrm{d}y$. Pomocí Rieszovy věty o reprezentaci se totiž dá dokázat, že RKHS jsou jednoznačně určeny *jádry*, tj. symetrickými pozitivně semidefinitními funkcemi $K\colon \Omega \times \Omega \to \mathbb{R}$, kde Ω je libovolná neprázdná množina. Výklad teorie prostorů s reprodukčním jádrem přesahuje rámec tohoto článku, podrobně je popsána v knihách [2], [21], stručnější přehledy vlastností RKHS lze nalézt např. v [5], [18].

Pro ilustraci, jakou roli hrají normy na těchto prostorech, uvedeme následující příklad: Je-li K konvoluční jádro, tj. K lze vyjádřit jako $K(x,y) = k(x-y)$, kde $k\colon \mathbb{R} \to \mathbb{R}$, a jestliže k má pozitivní Fourierovu transformaci, potom norma na RKHS určeném K označovaná $\|\cdot\|_K$ má tvar

$$\|f\|_K^2 = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \frac{\tilde{f}(\omega)^2}{\tilde{k}(\omega)}\,\mathrm{d}\omega.$$

To znamená, že pro konvoluční jádra s pozitivní Fourierovou transformací tvoří normy na RKHS vysokofrekvenční filtry. Paradigmatický příklad jádra je konvoluční jádro definované pomocí Gaussovy funkce, jejíž Fourierova transformace je $\exp(-\|\omega\|^2/2)$.

Pomocí teorie pseudoinverze aplikované na operátory L_x se dají odvodit vlastnosti řešení úlohy minimalizace funkcionálu empirické chyby $\mathcal{E}_z$ na RKHS. Pro jednoduchost popíšeme jen případ, kdy jádro splňuje silnější podmínku pozitivní definitnosti místo semidefinitnosti [2]. Pro tato jádra má řešení tvar

$$f^+(x) = \sum_{i=1}^{m} c_i K(x_i, x)$$

s vektorem koeficientů $c = (c_1, \dots, c_m)$ splňujícím $c = \mathcal{K}[x]^+ y$, kde $\mathcal{K}[x]$ je *Gramova matice jádra K vzhledem k vektoru vstupních dat* $x = (x_1, \dots, x_m)$, definovaná jako $\mathcal{K}[x]_{i,j} = K(x_i, x_j)$, $i, j = 1, \dots, m$, viz [10]. To znamená, že řešení se dá spočítat pomocí Mooreovy-Penroseovy pseudoinverze matice $\mathcal{K}[x]$ aplikované na vektor výstupních dat $y = (y_1, \dots, y_m)$. Řešení interpoluje data, tj. minimum $\mathcal{E}_z$ je rovno nule. Jeho stabilita vzhledem k malé změně vektoru y závisí na čísle podmíněnosti $\mathcal{K}[x]$, které je pro pozitivně definitní matice rovno podílu největšího a nejmenšího vlastního čísla.

4. Generalizace modelovaná pomocí regularizace

Globální podmínka modelující schopnost generalizace může být ještě zesílena tím, že kromě toho, že hledáme řešení pouze mezi funkcemi s konečnou K-normou, navíc ještě penalizujeme řešení, která mají tuto normu velkou. Tichonovova regularizace [20] nahrazuje problém minimalizace funkcionálu $\|A(\,\cdot\,) - g\|_Y^2$ na prostoru X hledáním funkce f^γ splňující rovnici

$$\|A(f^\gamma) - g\|_Y^2 + \gamma\|f^\gamma\|_X^2 = \min_{f \in X} \|A(f) - g\|_Y^2 + \gamma\|f\|_X^2.$$

Pomocí regularizačního parametru γ lze nastavit míru této penalizace řešení s velkou normou $\|\cdot\|_X$.

Minimalizace hodnoty funkcionálu $\mathcal{E}_z(f)+\gamma\|f\|_K^2 = \|L_x(f)-y/\sqrt{m}\|_2^2+\gamma\|f\|_K^2$ modeluje úlohu hledání funkce, která nemá příliš velké oscilace typu daného volbou jádra K a současně dobře aproximuje data. Dá se ukázat, že tato úloha má jediné řešení tvaru

$$f^\gamma(x)=\sum_{i=1}^m c_i^\gamma K(x_i,x),$$

kde $c^\gamma=(\mathcal{K}[x]+\gamma m\mathcal{I})^{-1}y$ a $\mathcal{I}$ značí jednotkovou matici (viz [5], [18], [10]).

V regularizovaném i neregularizovaném případě je tedy ideálním řešením úlohy učení neuronová síť, která má stejný počet výpočetních jednotek ve skryté vrstvě, jako je velikost trénovací množiny $\{(x_i,y_i)\colon i=1,\dots,m\}$, a jejíž výpočetní jednotky počítají jádrové funkce s parametry danými vstupními daty (v případě radiálních sítí se parametry nazývají středy). Řešení regularizované i neregularizované úlohy se liší pouze koeficienty lineární kombinace výpočetních jednotek, které odpovídají výstupním vahám sítě. V neregularizovaném případě platí $c=\mathcal{K}[x]^+y$, vektor výstupních vah c je tedy roven Mooreově-Penroseově pseudoinverzi Gramovy matice jádra K vzhledem k vektoru vstupních dat x aplikované na vektor výstupních dat y. V regularizovaném případě platí $c^\gamma=(\mathcal{K}[x]+\gamma m\mathcal{I})^{-1}y$, vektor výstupních vah c^γ je tedy roven inverzi matice $\mathcal{K}[x]+\gamma m\mathcal{I}$ aplikované na vektor y.

Popsané řešení problému učení na základě empirických a konceptuálních dat modelovaných pomocí norem na RKHS lze využít k návrhu algoritmů učení neuronových sítí založených na řešení soustav lineárních rovnic $c=\mathcal{K}[x]^+y$ a $c^\gamma=(\mathcal{K}[x]+\gamma m\mathcal{I})^{-1}y$. Pro gaussovské $K(x,y)=\exp(\|x-y\|^2)$ jádro konstruují tyto algoritmy radiální sítě, které jsou hojně používaným modelem neuronové sítě.

Matematická teorie učení tedy nabízí nové algoritmy jako alternativu ke standardním algoritmům učení využívajícím gradientní nebo evoluční metody. Porovnání výhod a i nevýhod obou typů algoritmů je předmětem výzkumu (viz [11], [12], [9]). Výhodou algoritmů založených na řešení popsaných soustav lineárních rovnic je, že vedou k nejlepším možným řešením úlohy učení. Meze jejich využití jsou však dány časovou náročností řešení soustav lineárních rovnic potřebných pro výpočet koeficientů odpovídajících výstupním vahám sítí a možnou nestabilitou těchto koeficientů vzhledem k šumu.

Zlepšení stability pomocí regularizace můžeme odhadnout porovnáním čísel podmíněnosti matic $\mathcal{K}[x]$ a $\gamma m\mathcal{I}+\mathcal{K}[x]$:

$$\operatorname{cond}(\mathcal{K}[x]+\gamma m\mathcal{I})=1+\frac{\operatorname{cond}(\mathcal{K}[x]-1)\lambda_{\min}}{\lambda_{\min}+\gamma m}\,,$$

kde $\lambda_{\min}$ značí nejmenší vlastní číslo matice $\mathcal{K}[x]$. Zvyšováním regularizačního parametru γ se tedy zlepšuje stabilita, ale současně se zhoršuje aproximace empirických dat.

Literatura

[1] Aronszajn, N.: *Theory of reproducing kernels.* Trans. Amer. Math. Soc. *68* (1950), 33–404.

[2] Berg, C., Christensen, J. P. R., Ressel, P.: *Harmonic Analysis on Semigroups.* Springer-Verlag, New York 1984.

[3] Bertero, M.: *Linear inverse and ill-posed problems.* Advances in Electronics and Electron Physics *75* (1989), 1–120.

[4] Bjorck, A.: *Numerical methods for least squares problem.* SIAM 1996.

[5] Cucker, F., Smale, S.: *On the mathematical foundations of learning.* Bull. Amer. Math. Soc. *39* (2001), 1–49.

[6] Friedman, A.: *Modern Analysis.* Dover, New York 1982.

[7] Girosi, F.: *An equivalence between sparse approximation and support vector machines.* Neural Computation *10* (1998), 1455–1480 (AI Memo No 1606, MIT).

[8] Groetch, C. W.: *Generalized Inverses of Linear Operators.* Dekker, New York 1977.

[9] Kůrková, V.: *High-dimensional approximation by neural networks.* Chapter 4 in *Advances in Learning Theory: Methods, Models and Applications* (J. Stuykens et al., ed.) (2003), 69–88. IOS Press, Amsterdam.

[10] Kůrková, V.: *Learning from data as an inverse problem.* In Proc. of COMPSTAT 2004 (J. Antoch, ed.), Physica-Verlag, Heidelberg, 1377–1384.

[11] Kůrková, V., Sanguineti, M.: *Error estimates for approximate optimization by the extended Ritz method.* SIAM J. Optim. (to appear).

[12] Kůrková, V., Sanguineti, M.: *Learning with generalization capability by kernel methods with bounded complexity.* J. Compl. (to appear).

[13] Moore, E. H.: *Abstract.* Bulletin AMS *26* (1920), 394–395.

[14] Narcowich, F. J., Sivakumar, N., Ward, J. D.: *On condition numbers associated with radial-function interpolation.* J. Math. Anal. Appl. *186* (1994), 457–485.

[15] Parzen, E.: *An approach to time series analysis.* Annals Math. Statistics *32* (1966), 951–989.

[16] Penrose, R.: *A generalized inverse for matrices.* Proc. Cambridge Philos. Soc. *52* (1955), 406–413.

[17] Poggio, T., Girosi, F.: *Networks for approximation and learning.* Proc. IEEE *78* (1990), 1481–1497.

[18] Poggio, T., Smale, S.: *The mathematics of learning: dealing with data.* Notices Amer. Math. Soc. *50* (2003), 536–544.

[19] Sejnowski, T. J., Rosenberg, C.: *Parallel networks that learn to pronounce English text.* Complex Systems *1* (1987), 145–168.

[20] Tichonov, A. N., Arsenin, V. Y.: *Solutions of Ill-posed Problems.* W. H. Winston, Washington, D. C. 1977.

[21] Wahba, G.: *Splines Models for Observational Data.* SIAM, Philadelphia 1990.

[22] Werbos, P. J.: *Backpropagation: Basics and New Developments.* The Handbook of Brain Theory and Neural Networks (M. Arbib, ed.), 134–139. MIT Press, Cambridge 1995.