Petr Raab
Věty o univerzalitě a konzistenci neuronových sítí
Posudek oponenta diplomové práce

V práci jsou představeny základní postupy odhadu regresního modelu pomocí neuronových sítí. V první kapitole se čtenář seznámí se samotnou myšlenkou neuronových sítí a jejich budování. Druhá kapitola se věnuje problematice uiverzality, tedy schopnosti neuronových sítí modelovat či blízce aproximovat funkce z velmi široké třídy. Ve třetí kapitole se autor zabývá asymptotickými vlastnostmi odhadu pomocí neuronových sítí, tedy konzistenci a asymptotické normalitě. Poslední kapitola obsahuje simulační studii, ve které jsou porovnávány výsledky mělkých neuronových sítí pro čtyři různé regresní funkce, z toho dvě funkce jedné proměnné a dvě funkce dvou proměnných, a čtyři rozsahy výběrů od 500 do 5 000.

Obsahově je práce budovaná logicky a obsahuje v zásadě vše podstatné pro pochopení základů neuronových sítí a jejich nejdůležitějších vlastností. Na druhou stranu jazyková stránka práce značně pokulhává za obsahovou. Kromě mnohdy nešikovných fomrulací jsou v práci i časté překlepy a gramatické chyby. To pak zbytečně snižuje jinak dobrý dojem z práce a má první výtka tak směřuje k doporučení příště věnovat více času závěrečným úpravám textu.

Kromě několika středně závažných a drobných výtek bych během obhajoby rád diskutoval otázku, co vlastně neuronové sítě přinášejí navíc proti ostatním známým metodám.

- Máme-li vektor vysvětlujících proměnných jen o několika souřadnicích, dá neuronová síť lepší výsledky než běžné lokální odhady regresní funkce? Pokud ano, za jakou cenu?
- Má-li vektor vysvětlujících proměnných vysokou dimenzi (problém $p > n$, je neuronová síť schopná označit, které vysvětlující proměnné jsou v modelu podstatné, podobně jako například LASSO metoda?
- Dá se neuronová síť považovat za svého druhu metodu redukce dimenze? Bylo by možné průběžné výsledky, obvykle nazývané vrstvy dopředné neuronové sítě, použít místo jiných metod redukce dimenze, například metody hlavních komponent?
- Pokud na některou z těchto otázek je odpověď ano, pak je škoda, že v práci takové srovnání nebylo vyzkoušeno.

K práci mám i několik dalších dotazů a připomínek, které bych rád probral během obhajoby.

(1) Značení prostoru $1 \times \mathbb{R}^n$ je nestandardní a mělo by být vysvětleno, co znamená kartézský součin čísla a množiny.
(2) V prvním odstavci části 1.3 je naznačeno, že k dalším výpočtům se bude hodit diferencovatelnost aktivačních funkcí a hned ve druhém odstavci je nejjednodušší aktivační funkce, která je v nule nediferencovatelná.
(3) Na konci části 1.3 je zmíněna klesající a zároveň nemonotónní funkce. Jak si takovou funkci máme představit?
(4) Na úplném konci první kapitoly se konstatuje, že rychlost růstu odhadovaných parametrů je velmi rychlá. Divergují tedy odhadované parametry do nekonečna, nebo mají nějakou horní asymptotu?
(5) Ve druhém odstavci části 2.1 jsou zmíněné exponenciální aproximační schopnosti. Odhlédneme-li od chybných pádů v této větě (dosahují *čeho*, ne *čemu*), co tyto schppnosti představují?

(6) V souvislosti s definicí 6 a větou hned za ní, jaký je rozdíl mezi afinní a lineární funkcí?
(7) V definici 9 je množina S označená za *stejnoměrně hustou na kompaktu v* $\mathcal{C}(\mathbb{R}^{q_0})$. Zde není úplně jasné, čeho se týká slovo kompakt, zda jde o podmnožinu prostoru spojitých funkcí, nebo zda jde o podmnožinu nosiče těchto funkcí.
(8) Ve větě 2 je předpoklad na aktivační funkci velmi slabý, má jít o měřitelnou a neklesající funkci s limitami v nekonečnu. Lze nějak vysvětlit, proč tato věta platí například s volbou $\phi(x) = \mathbf{1}(x \geq 0)$?
(9) Druhý odstaven na straně 17 obsahuje nějaký překlep, ale není snadné zjistit, jaké je správné znění. Pokud $G_\epsilon(\lambda) \in [0, \epsilon/2Q)$ a $G_\epsilon(\lambda) \in [0, 1/Q]$, pak opravdu jejich rozdíl je mezi nulou a maximem z $\epsilon/2Q$ a $1/Q$, protože je nulový. To ale asi není tvrzení, které zde má být.
(10) Stále na straně 17 uprostřed je věta začínající slovy „Musíme si uvědomit …“. Dal bych zde přednost důkazu, protože nahlédnout zde uvedenou skutečnost není snadné.
(11) Pokud všude ve třetí kapitole požadujeme nezávislost ϵ_i na $\mathbf{X}_i$, nestačí pak předpokládat pouze $\mathsf{E}\epsilon_i = 0$ a $\mathsf{E}|\epsilon_i|^{2+\delta} < \infty$?
(12) Ve větě 5 jsou zavedené funkce $M_n(\theta)$ a $M(\theta)$ co je definičním oborem a oborem hodnot těchto funkcí? Jedná se o náhodné funkce (soudě podle požadavku na konvergenci v pravděpodobnosti)?
(13) Na řádcích 2 a 3 na straně 25 jsou dvě ostré nerovnosti. Zdá se mi, že jsou ve sporu.
(14) Na straně 27 nahoře je argumentováno, že funkce má omezenou derivaci, protože její derivace v plus i mínus nekonečnu jsou nulové. Opravdu taková implikace platí?
(15) Proč v simulační studii není zahrnut případ, kdy je nutné odhadnout σ^2?

Chci zdůraznit, že výše uvedené problémy nesjou zásadní a dají se snadno vysvětlit a vyřešit. Je však škoda, že tyto nedostatky, a je jich v práci více, zbytečně snižují čitelnost a celkové vyznění práce.

Přes všechny tyto výhrady se domnívám, že práce splnila svůj účel a je dostatečně dobrá na to, aby mohla při obhajobě být **uznána za diplomovou práci**.

Ve Zbraslavi nad Vltavou dne 4. června 2024

Daniel Hlubinka