Posudek oponenta bakalářské práce

Název: Stejnoměrný zákon velkých čísel, VC dimenze a strojové učení
Autor: Aibat Kossumov

Shrnutí obsahu práce

Autor nejprve studuje *Glivenkovu-Cantelliho větu* a *VC dimenzi* tříd funkcí a množin. Část o Glivenkově-Cantelliho větě obsahuje její podrobný důkaz pomocí tzv. *pokrývacích čísel* a je zde ukázáno, že tato věta zobecňuje klasický výsledek ohledně konvergence empirické distribuční funkce. V části o VC dimenzi, která popisuje složitost třídy funkcí (či množin), je dokázán vztah mezi VC dimenzí a *Glivenkovu-Cantelliho* vlastností. Oba pojmy jsou následně porovnány v kontextu základní věty statistického učení. Autor ukazuje, že základní věta statistického učení v některých klasifikačních úlohách lze ekvivalentně vyjádřit pomocí výše uvedeného. Tato klasická věta udává charakterizaci úloh, které se „dají učit“. Celý text je doplněn názornými příklady.

Celkové hodnocení práce

Téma práce. Téma je náročné, neboť je nutné velmi obezřetně pracovat s neměřitelnými zobrazeními a zobecněnou konvergencí. Student toto zvládl výborně. I proto se dá považovat téma za zvládnuté.

Vlastní příspěvek. Práce je zejména rešeršní, ovšem přehledné rozepsání některých velmi technických důkazů společně s uceleným popisem kontextu je velmi hodnotným příspěvkem.

Matematická úroveň. V práci se nevyskytují vážnější chyby ve formulacích matematických tvrzení ani jejich důkazech. Toto s ohledem na fakt, že při manipulaci se zobecněnou konvergencí skoro jistě a v pravděpodobnosti musí být matematik velmi obezřetný, ukazuje autorovu výbornou matematickou úroveň.

Práce se zdroji. Zdroje, ze kterých student vychází, jsou uvedeny. Náznak plagiátorství není z textu patrný.

Formální úprava. Formální úprava odpovídá úrovni bakalářské práce. Práce obsahuje některé jazykové nedostatky, avšak dobrá čitelnost textu je zachována.

Otázky

1. Pokrývací čísla zdánlivě souvisí s tzv. ϵ-sítí a totální omezeností metrického prostoru. Jaký je vztah těchto dvou pojmů? Pro určitost můžeme uvažovat supremální metriku na prostoru omezených funkcí.

2. Na str. 11 uvádíte argument, proč v definici pokrývacích čísel uvažujeme pouze koule se středem s konečnou normou $||\cdot||$. Tento argument se zdá jako nepřesvědčivý (minimálně neplatí, že z nerovností v první odsazené formuli na str. 11 vyplývá, že $||f|| > \infty$). Existuje jiný důvod, proč neuvažovat středy koulí o nekonečné normě?

3. V Poznámce 3.1 zmiňujete, že v této práci nerozlišujete pojmy *PAC learnable* a *PAC agnostic learnable*. Můžete popsat, jaký je v těchto pojmech rozdíl? Který z těchto pojmů je zajímavější v konkrétních aplikacích?

4. V Poznámce 3.5 správně uvádíte, že ačkoliv v důkazech s výhodou využíváme algoritmus, který minimalizuje empirické riziko, toto může být v praxi problematické. Jako příklad uvádíte tzv. *overfitting.* Jaký další problém s minimalizací empirického rizika může nastat? Vezměte v potaz, že v praxi tento problém často vede na minimalizaci funkce, která není konvexní.

5. V Lemma 1.3 pracujeme s náhodným zobrazením $||P_n^0||_{\mathcal{F}}$, což je funkce definována na $(\Omega \times \Omega_\epsilon)$. Jak máme chápat výraz $P^* \left(||P_n^0||_{\mathcal{F}} > \frac{\eta}{4}\right)$, když P je pravděpodobnostní míra na $(\Omega, \mathcal{A})$? Je P opravdu předmětem předvedeného důkazu?

6. V důkazu Věty 1.6 na str. 16 uvádíte, že $|g_j^\omega| \leq M$. Jak máme tuto nerovnost chápat, když g_j^ω je funkce?

7. V Příkladu 3.1 používáte označení J pro nějakou množinu, ačkoliv toto značení není uvedeno. Jaký je význam/využití faktu, že $\hat{I}_\omega \subset J$?

8. V příkladu 3.1 v inkluzi $\{X_i(\omega)...\} \subset J$ užíváte (3.3) v silnější formě nejen pro $i = 1$, ale pro každé $i = 1, \ldots, n$. Dá se tato zesílená varianta odvodit již z Vaší varianty (3.3) s pomocí dalších předpokladů, které máte?

9. V příkladu 3.1 předpokládáte (3.4). Je potřeba (3.4) předpokládat, je-li dáno (3.3)? V další větě nesprávně aplikujete negaci (3.4), neboť (3.4) závisí pouze na jedné zvolené pravděpodobnosti P, kterou jste zvolil výše. Tedy negace by měla hovořit pouze o tomto P (zde je rozdíl oproti variantě, kdy by (3.4) obsahovala existenční kvantifikátor k P).

10. V důkazu Věty 3.1 tvrdíte, že $L_{P_i}(h_n^\omega)$ je měřitelné, avšak ve (3.7) jste okomentoval měřitelnost pouze indikátoru. Jak je to s měřitelností $L_{P_i}\left(ALG(S_j^i)\right)$?

Připomínky

1. Důkaz Lemma 1.3 obsahuje formuli (1.1), ve které první z nerovností neplatí. Chyba se dá přímočaře opravit, pokud uvažujeme $n \geq \frac{8M^2}{\eta^2}$, což bylo koneckonců předpokládáno.
2. V popisu příkladu se dvěma hráči na str. 29 postrádáme definici y_i.
3. Jazykové nepřesnosti se vyskytují v rámci celého textu. Nepřesnost v matematických formulích byla nalezena pouze v důkaze Lemma 1.3, kdy je dvakrát nesprávně uzávorkováno.

Závěr

Práci považuji za vynikající a doporučuji ji uznat jako bakalářskou práci.

Mgr. Ondřej Týbl
Katedra pravděpodobnosti a matematické statistiky MFF UK
6.6.2022