

POSUDEK OPONENTA BAKALÁŘSKÉ PRÁCE

Název: Generativní modely dat popsaných stromovou strukturou

Autor: Jakub Bureš

SHRNUTÍ OBSAHU PRÁCE

Práce se zabývá Bayesovskou analýzou generativních modelů a jejich aplikací na hierarchické modely strojového učení. Ty tvoří speciální oblast strojového učení, neboť jednotlivé vzorky nemusí obsahovat stejný počet featur, což tento typ modelů dělá obtížnějším. První kapitola práce je teoretický úvod, ve kterém je velké množství chyb. Dokonce mi přišlo, že v mnoha částech se autor snažil něco vysvětlit, ale protože ne zcela rozuměl problematice, text nenapsal dobře. Druhá a třetí kapitola obsahují aplikaci teoretické části na generativní modely a na stromové (hierarchické) struktury. I tato část obsahuje značné množství chyb. Některé části nejsou dobře vysvětleny a návaznost mezi jednotlivými částmi mi nepřijde dostatečně vysvětlena. V některých případech se stalo, že mezikrok je špatně, ale výsledek dobře. Značení není jednotné, například w a θ se používá pro to samé; podobně n označuje jak počet vzorků, tak jejich dimenzi.

Autorovi bych doporučoval prohloubit si matematické základy, které mu dle mého názoru docela chybí. Zároveň bych doporučil věci psát pořádně, protože mi přišlo, že spousta chyb pramení ze špatného zápisu.

CELKOVÉ HODNOCENÍ PRÁCE

Práce je svým obsahem velmi náročná a spíše než jako bakalářská práce mi přijde jako magisterská. Autor se se složitým tématem popral dle mého názoru dobře. I přes spoustu věcí, které práci vytýkám, se mi práce obsahově líbí a *doporučuju ji uznat jako práci bakalářskou*. Pokud existuje možnost, preferoval bych vypracování errat. Znamku navrhu podle průběhu obhajoby.

POZNÁMKY K PRÁCI

1. Autor uvádí, že SGD je vylepšenou verzí GD. Toto není úplně pravda. Kdy a proč se používá SGC místo GD?
2. Je uvedeno, že ADAM, RMSProp jsou verze SGD. To není pravda. Mohou být použity i v kombinaci s GD?
3. Celá první kapitola předpokládá, že vstupy do regrese jsou jednodimenzionální.
4. Analytické řešení ze strany 8 předpokládá, že je perfektní fit (nulová chyba). Jak se dostane odhad bez tohoto předpokladu?
5. V Definici 1.2.1 třetí podmínka není dobře.
6. Ve Větě 1.2.1 není dobře druhý bod.
7. Věta 1.2.4 není věta ale definice. Navíc si myslím, že není dobře pro více než dvě množiny.
8. Definice 1.2.3 je špatně. Funkce zobrazuje pouze do \mathbb{R}^n a ne do součinu s Borelovskou σ -algebrou. Ta tam pouze vystupuje tak, že funkce musí být měřitelná.
9. Definice 1.2.4 je špatně. Hustota nemusí být spojitá ani pro absolutně spojitá náhodné veličiny.
10. Co je regulární zobrazení ve Větě 1.2.5?

11. V (1.18) je divné značení, které mi nepříjde dobře.
12. U inverzního gamma rozdělení je špatně rozptýl.
13. Poznámku za inverzním gamma rozdělení nechápu. Daná funkce se nenaintegruje na jedničku, takže podle mně není hustota.
14. Pro (1.28) je nutné předpokládat, že každá složka má normální rozdělení a nejenom jedna.
15. Co je ve vztahu (1.34) α ? Pokud jsem nic nepřehlédl, tak se předtím v práci nevyskytovalo.
16. Poznámka pod definicí divergence je špatně. Semimetrika splňuje symetrii. Jaký je rozdíl mezi metrikou a semimetrikou?
17. Vzorec (1.41) je špatně. Na jeho základě byl odvozen (1.42), který je ale už dobře. Přijde mi, že autor spíše opisoval vzorečky, než aby se nad nimi zamyslel.
18. V odstavci pro Kullback-Leibler divergenci autor tvrdí, že logaritmus je konvexní funkce. To není úplně pravda.
19. Druhá rovnost v (1.47) je špatně. Výsledek v (1.48) je ale opět dobře.
20. Druhá rovnost v (1.49) je špatně.
21. Poslední rovnost v (1.51) nevidím. Jak se rozepíše?
22. V (1.54) se najednou objeví 4 další proměnné. O co se jedná?
23. V Definici 1.3.2 je špatné značení.
24. Co se myslí vzorcem (2.4)? Levá strana je dvourozměrná veličina, zatímco pravá strana vypadá jako jednorozměrná.
25. Z Kapitoly 2.2 jsem zmaten. Zatímco v této kapitole se (x_1, \dots, x_n) používá jako jeden vzorek dimenze n , v předchozím textu se jednalo o n vzorků dimenze 1.
26. Vzorec (2.5) má špatnou sumu.
27. U vzorců (2.7)-(2.9) musí být použity horní indexy i pro z a a .
28. Zatímco y v (2.6) je jednorozměrné, v (2.10) je vícerozměrné. Proč je tomu tak?
29. Před (2.16) je zmíněna věrohodnostní funkce. Ta je ale s mínusem (aby se maximalizovala a ne minimalizovala věrohodnost).
30. V (2.16) jsou špatně indexy v prostředním řádku.
31. Od (2.18) do (2.24) není jasné, co se dělá. Operace jsou náročné a měly by se pořádně rozepsat.
32. Vzorec na posledním řádku před (3.3) neplatí.
33. Před (3.10) se píše, že je jasné, že se bude jednat o normální rozdělení. Proč je to jasné? Platí to obecně nebo za nějakých předpokladů? Například pro parametr λ ?

Lukáš Adam

Fakulta elektrotechnická, České vysoké učení technické v Praze, Praha

27. 7. 2020