

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Cluster separability in multidimensional biomedical data
Jméno autora:	Klikarová Ivana
Typ práce:	bakalářská
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra teorie obvodů
Oponent práce:	Ing. Michal Šimek
Pracoviště oponenta práce:	Katedra teorie obvodů

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	průměrně náročné
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Zadání požaduje seznámení se s dostupnými metodami pro estimaci optimálního počtu shluků pro shlukovací algoritmy a jejich aplikaci a vyhodnocení na simulovaných a reálných datech.	

Splnění zadání	splněno
<i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>	
Zadání bylo splněno.	

Zvolený postup řešení	správný
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Postup řešení považuji za správný.	

Odborná úroveň	D - uspokojivě
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Úvodní přehled metod je popsán poměrně stručně a některé algoritmy a statistické metody používané při analýze nejsou zmíněny vůbec (např. PCA, chí-kvadrát test). I když se v práci prakticky nevyskytují faktické chyby, některá tvrzení by bylo vhodné formulovat jinak pro zajištění konzistentnosti a technické správnosti textu. Získaný dataset obsahující data o pacientech s COVID-19 by si zasloužil detailnější a ucelenější přehled obsažených informací, včetně např. typu a rozsahu jednotlivých symptomů. V práci chybí popis implementace jednotlivých metod. Není jasné, jak přesně byly metody realizovány, což ztěžuje reprodukci výsledků a pochopení použitého přístupu. Postup analýzy reálných dat je sice popsán stručně, ale vyhodnocení výsledků je provedené správně a odpovídá standardním postupům.	

Formální a jazyková úroveň, rozsah práce	B - velmi dobře
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
K formální stránce mám několik drobných výhrad. V práci se vyskytuje několik překlepů, které však zásadně neovlivňují pochopitelnost textu. Titulky obrázků by měly být dostatečně podrobné, aby samy vysvětovaly jejich význam. Některé obrázky mají navíc české popisky. Nicméně, po grafické stránce je práce kvalitně zpracována. V úvodní části, která se věnuje rešerši dostupných metod, by bylo pro čtenáře přínosné sjednotit zápis matematických výrazů, konstant a množin, aby byl text čitelnější a typografie konzistentní (např. množina shluků nebo středy shluků (centroids) jsou v každé definici označeny jinak).	

Výběr zdrojů, korektnost citací	A - výborně
<i>Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně</i>	

odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Bibliografické citace jsou v souladu s citačními zvyklostmi a normami.

Další komentáře a hodnocení

Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.

Z práce není zřejmé, které metody byly implementovány studentkou, v jakém programovacím jazyce či s použitím jakých knihoven. Práce neobsahuje žádné přílohy, které by umožnily zjistit případnou implementaci.

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.

Studentka nejprve shrnuje dostupné metody používané k určení optimálního počtu shluků pro shlukovací algoritmy. Tyto metody byly poté aplikovány na simulovaná data s využitím algoritmu K-means. Následně jsou vybrané metody použity na reálná data pacientů s COVID-19. Práce navíc obsahuje analýzu těchto dat, zahrnující explorační analýzu, předzpracování, redukci dimenze pomocí PCA a statistickou analýzu s cílem interpretovat informace zachycené v hlavních komponentách a určit charakteristiky skupin, které byly klasifikačním algoritmem získány. Jako hlavní nedostatek práce vnímám chybějící popis implementace jednotlivých metod a konfigurace jejich parametrů.

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **C - dobře**.

Otázky:

- 1) Jaké knihovny byly použity při implementaci optimalizačních metod? Implementovala si studentka některé metody sama?
- 2) Z výsledků vyplývá, že metoda "Elbow", která se často používá pro prvotní odhad počtu shluků selhává i při analýze nízkošumových, dobře separovatelných dat. Jaký práh (threshold) byl nastaven pro určení optimálního počtu clusterů?
- 3) Jakým způsobem byla použita křížová validace (cross-validation) pro odhad optimálního počtu shluků?

Datum: 31.5.2024

Podpis: