

1. úloha „Analýza a využití klinických dat“

Zadání

1. Na základě analýzy dat stanovte rizikové faktory (resp. klasifikátor) pro následující výstupy:
 - Císařský řez
 - pH arteriální krve < 7.1

Zpracování a hodnocení

Analýza a využití dat v sobě skloubí statistiku a metody strojového učení. Ve skutečnosti je v mnoha ohledech statistika a strojové učení v překryvu (regrese). Vy v rámci této úlohy provedete základní statistickou analýzu a následně si můžete zvolit jeden ze dvou postupů pro další komplexnější zpracování dat. To znamená, že při vypracování obou přístupů komplexnějšího zpracování se body nad minimální počet počítají jako zpracování rozšířené části. Tady pozor! Nelze kombinovat body z obou přístupů pro získání minimálního počtu bodů, tedy pro splnění úlohy musíte dosáhnout 3 bodů u jednoho ze dvou přístupů.

Úloha obsahuje ještě třetí část, která je vždy brána jako rozšiřující a nepočítá se do základní části úlohy.

Povinná společná část

1. Proved'te četnostní analýzu všech parametrů (číselně i graficky) a četnostní analýzu všech parametrů na základě výstupů (opět číselně i graficky). [1 bod]
Konzultace v průběhu zadávání úlohy:
 -
2. Proved'te vhodné testy pro určení signifikance jednotlivých parametrů vzhledem k rozdělení dle výstupních veličin (stačí číselné vyjádření). [1 bod]
Konzultace v průběhu zadávání úlohy:
 -
3. Proved'te univariační analýzu pro určení rizikových faktorů (např. Odds Ratio) zvlášť ve vztahu ke každému z výstupů (stačí číselné vyjádření). [1 bod]
Konzultace v průběhu zadávání úlohy:
 -

Povinná volitelná část

4. **První možnost:** Proved'te multivariační analýzu dat s využitím logistické regrese. Na základě této analýzy určete rizikové kombinace faktorů pro jednotlivých výstupy (stačí číselné vyjádření pomocí všech vhodných kritérií). [2 bod]
Konzultace v průběhu zadávání úlohy:
 -

Druhá možnost: Vytvořte klasifikátory, které bude na základě vybrané části dat určovat příslušnost k jednotlivým výstupům a klasifikaci validujte (stačí číselné vyjádření pomocí všech vhodných kritérií). [2 bod]

Konzultace v průběhu zadávání úlohy:

•

5. **První možnost:** Převeďte úlohu multivariační analýzy na úlohu klasifikace. Tuto klasifikaci validujte (stačí číselné vyjádření pomocí všech vhodných kritérií). [1 bod]

Konzultace v průběhu zadávání úlohy:

•

Druhá možnost: Získejte, na základě volby dat pro konstrukci klasifikátoru, nejlepší možný výsledek (stačí číselné vyjádření pomocí všech vhodných kritérií). [1 bod]

Konzultace v průběhu zadávání úlohy:

•

Nepovinná část

6. Doplněte do úlohy u bodů 2, 3 a 4 grafickou reprezentaci výsledků. [1 bod]

Konzultace v průběhu zadávání úlohy:

•

7. Proveďte selekci faktorů některou z metod užívaných ve strojovém učení (např. na základě InfoGain) a porovnejte výsledky s výsledky analýzy signifikace, univariační analýzy a multivariační analýzy, resp. s výběrem příznaků klasifikátoru. [2 bod]

Konzultace v průběhu zadávání úlohy:

•

8. Proveďte výpočet podmíněné pravděpodobnosti pro nalezené rizikové faktory u jednotlivých výstupů. [1 bod]

Konzultace v průběhu zadávání úlohy:

•

9. Proveďte výpočet podmíněné pravděpodobnosti pro kombinaci nalezených rizikových faktorů u jednotlivých výstupů. [2 bod]

Konzultace v průběhu zadávání úlohy:

•

Odevzdávat budete archive (dle možností upload systému) s velmi dobře komentovaným kódem (může být v češtině) a možností jednoduše měnit vstupy jednotlivých analýz. Výstupem, pak budou výpisy výsledků a vykreslené grafy.

Na závěr kódu přidáte do komentáře cca 3 odstavce, které shrnou vaše nejdůležitější zjištění.

Termín odevzdání úlohy je pondělí 5. 11. 2018 23:59.