

# Otázky na skúšku z predmetu “Analýza zhlučkov a klasifikácia dát (2-PMS-116)”

Radoslav Harman, KAMŠ, FMFI UK

14. mája 2025

1. Uveďte jeden zmysluplný príklad aplikácie nehierarchickej (partičnej) analýzy zhlučkov. Definujte problém k-means a opíšte Lloydov algoritmus na jeho približné vyriešenie. Porovnajte metódu k-means s metódou k-medoids (uveďte vzájomné výhody a nevýhody týchto dvoch metód). Uveďte výhody a nevýhody týchto metód v porovnaní s inými metódami analýzy zhlučkov.
2. Uveďte jeden zmysluplný príklad aplikácie nehierarchickej (partičnej) analýzy zhlučkov. Vysvetlite DBSCAN princíp analýzy zhlučkov a voľne opíšte zodpovedajúci algoritmus. Uveďte výhody a nevýhody tejto metódy v porovnaní s inými metódami analýzy zhlučkov.
3. Uveďte jeden zmysluplný príklad aplikácie nehierarchickej (partičnej) analýzy zhlučkov. Vysvetlite princíp analýzy zhlučkov založenej na normálnom modeli. Uveďte optimalizačný problém, ktorý rieši analýza zhlučkov založená na normálnom modeli. Uveďte výhody a nevýhody zhlučkovania založeného na normálnom modeli v porovnaní s inými metódami analýzy zhlučkov.
4. Uveďte jeden zmysluplný príklad aplikácie hierarchickej analýzy zhlučkov. Porovnajte situácie, v ktorých je vhodnejšie použiť hierarchické zhlučkovanie než partičné zhlučkovanie. Opíšte význam, princíp a varianty aglomeratívneho hierarchického zhlučkovania.
5. Uveďte jeden zmysluplný príklad aplikácie hierarchickej analýzy zhlučkov. Porovnajte situácie, v ktorých je vhodnejšie použiť hierarchické zhlučkovanie než partičné zhlučkovanie. Opíšte význam a princíp divizívneho hierarchického zhlučkovania.
6. Opíšte akými metódami a indexami meriame presnosť (chybovosť) binárneho klasifikátora. Opíšte základné metódy odhadovania chyby klasifikácie (resubstitučná metóda a varianty krížovej validácie). Aké iné kvality klasifikátora potrebujeme zohľadňovať okrem jeho presnosti?

7. Odvoďte Bayesovo rozhodovacie pravidlo pre klasifikáciu do dvoch skupín (pre všeobecné  $p_1, p_2, C(1|2), C(2|1), q_1, q_2$ ). Napíšte predpis pre Bayesovo rozhodovacie pravidlo pre všeobecný počet skupín. Opíšte interpretáciu konštánt a premenných v Bayesovom rozhodovacom pravidle (Bayesovom klasifikátore). Vysvetlite ako súvisí metóda k najbližších susedov s teoretickým Bayesovým klasifikátorom.
8. Napíšte odvodenie binárneho lineárneho diskriminačného pravidla a (stručne) binárneho kvadratického diskriminačného pravidla z Bayesovho rozhodovacieho pravidla. Uveďte výhody a nevýhody takto skonštruovaných klasifikátorov v porovnaní s ostatnými najpoužívanejšími klasifikačnými metódami.
9. Vysvetlite čo je to klasifikačný strom a ako ho je možné skonštruovať metódou rekurzívneho delenia. Pritom definujte a zdôvodnite pojem "nečistoty uzla". Uveďte výhody a nevýhody klasifikačného stromu v porovnaní s ostatnými najpoužívanejšími klasifikačnými metódami.
10. Stručne vysvetlite pojmy osekávania (angl. *pruning*) stromu, metódu *bagging* konštrukcie klasifikátora a princíp konštrukcie náhodných lesov (angl. *random forests*). Diskutujte o výhodách a nevýhodách náhodných lesov v porovnaní s ostatnými najpoužívanejšími klasifikačnými metódami.
11. Stručne opíšte *boosting* slabých klasifikátorov a metódu Adaboost. Diskutujte o výhodách a nevýhodách klasifikátorov založených na boostingu v porovnaní s "jednoduchými" klasifikačnými metódami.
12. Opíšte základnú myšlienku metód oporných bodov pre prípad lineárne separovateľných dát a pre prípad lineárne neseparovateľných dát. Stručne vysvetlite základné myšlienky nelineárnych metód oporného bodu (napríklad vysvetlite pojem jadra). Uveďte základné výhody metód oporného bodu v porovnaní s ostatnými najpoužívanejšími klasifikačnými metódami.