

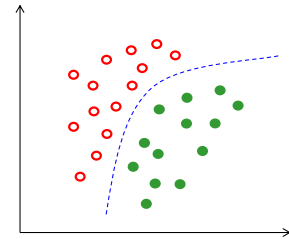
1. Jaké jsou předpoklady ANOVY pro vícerozměrná data?
  - a. Normalita v alespoň jedné skupině
  - b. Homogenita rozptylů všech srovnávaných skupin
  - c. Závislost skupin dat, které srovnáváme
2. U metrik podobnosti je podobnost objektu od sebe samého
  - a. Vždy hodnota 1
  - b. Maximální hodnota podobnosti pro danou metriku
  - c. Vždy hodnota 0
3. Metrika pro určení vzdálenosti mezi 2 objekty s kvalitativními proměnnými je
  - a. Hammingova metrika
  - b. Pearsonův korelační koeficient
  - c. Metoda nejbližšího souseda
4. Abychom vytvořili asociační matici mezi objekty, musíme provést
  - a. Q mode analýzu
  - b. O mode analýzu
  - c. R mode analýzu
5. Metody shlukové analýzy se dělí na
  - a. Jednoznačné a nejednoznačné
  - b. Aglomerativní a divizivní
  - c. Parametrické a neparametrické
6. Při redukci dat vybíráme proměnné
  - a. Které jsou spolu korelované
  - b. Které mají co největší vzdálenost mezi třídami
  - c. Tak, aby jich bylo co nejvíc
7. Při identifikaci optimálního počtu hlavních komponent u PCA využíváme
  - a. Sutinový graf
  - b. Dendrogram
  - c. Vícerozměrný t-test
8. Co **nepatří** do transformace vícerozměrných dat?
  - a. Standardizace
  - b. Min-max normalizace
  - c. Post hoc testy
9. Asociační matice vzdáleností má na diagonále
  - a. Vždy hodnoty 1
  - b. Vždy hodnoty 0
  - c. Nelze dopředu říci, záleží na konkrétních datech.
10. Pokud budeme chtít v budoucnu měřit méně proměnných, který typ redukce dat použijeme?
  - a. Selekcí
  - b. Extrakcí
  - c. Transdukci
11. Mezi metody hierarchického shlukování nepatří
  - a. Metoda k-průměrů
  - b. Centroidová metoda
  - c. Metoda průměrné vazby
12. Při klasifikaci 80 kontrolních subjektů a 20 pacientů byla dosažena úspěšnost 80%, což je:
  - a. Výborný výsledek
  - b. Neuspokojivý výsledek
  - c. Nelze zhodnotit bez znalosti vstupních dat

13. Umožňuje vždy první hlavní komponenta u PCA dobrou klasifikaci?

- Ano
- Ne
- Není známo

14. Která z uvedených metod je nejvhodnější pro klasifikaci dat zobrazených vpravo?

- Metoda podpůrných vektorů
- Analýza nezávislých komponent
- Lineární diskriminační analýza



15. Je resubstituce vhodným způsobem validace dat?

- Ano
- Ne
- Není známo

16. Která z uvedených metod umožňuje nelineární redukcí dat?

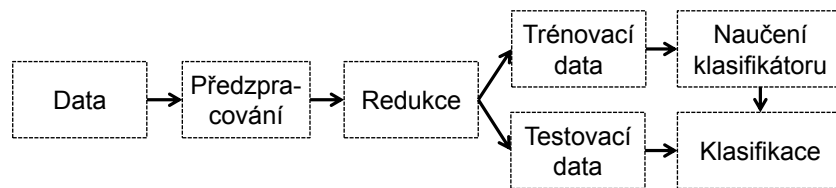
- Analýza hlavních komponent (PCA)
- Metoda lokálně lineárního vnoření (LLE)
- Vícerozměrné škálování (MDS)

17. Použití jaké z následujících asociačních matic vede k největší ztrátě informace u PCA?

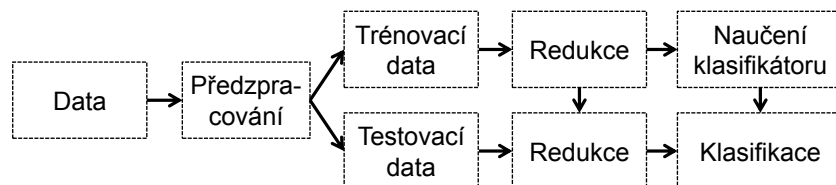
- Matice korelačních koeficientů (tzn. klasická korelační matice)
- Autokorelační matice
- Kovarianční matice

18. Jaký postup experimentu je správný?

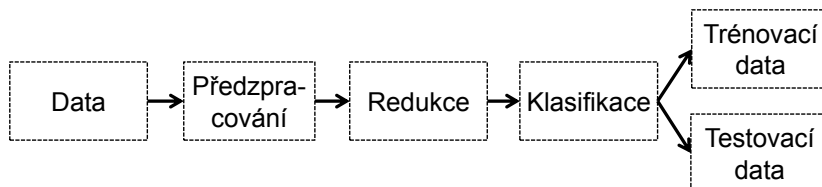
a.



b.



c.

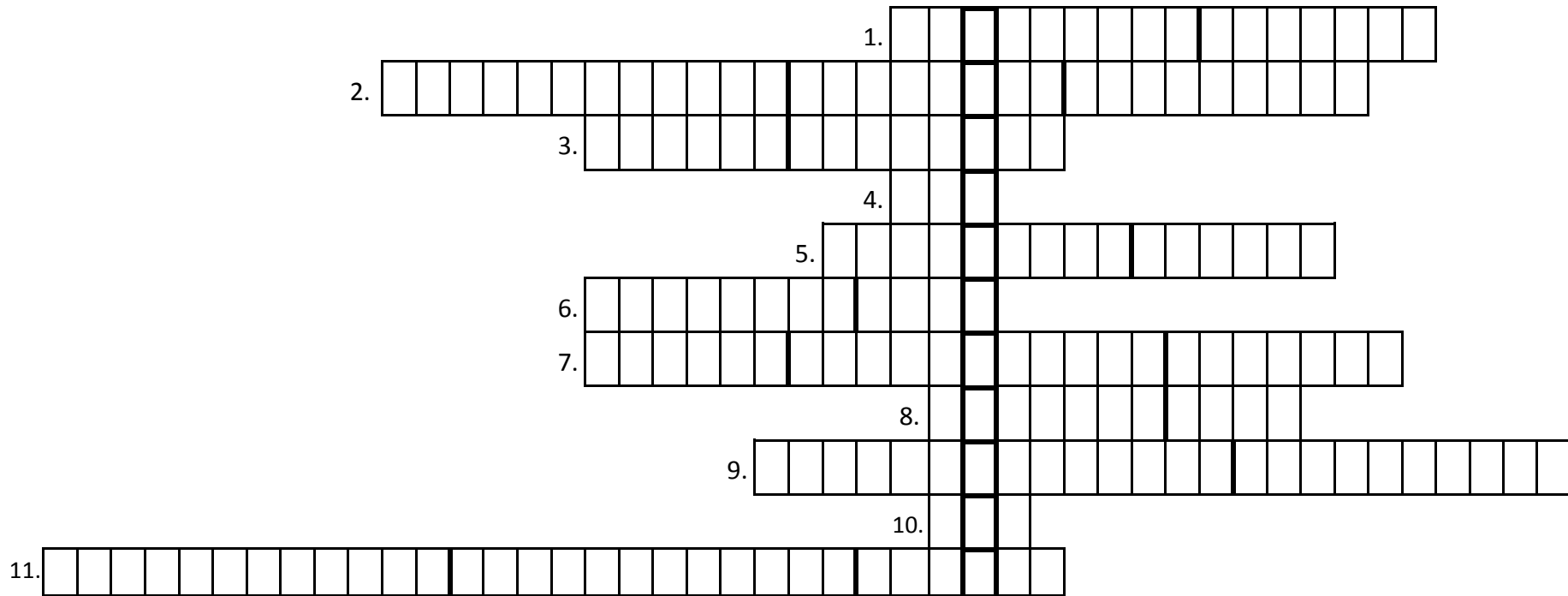


19. Kdy hrozí největší riziko nadhodnocených výsledků klasifikace při nesprávné validaci?

- Při výběru významných proměnných pomocí t-testu
- Při redukcí dat pomocí analýzy hlavních komponent
- Při redukcí dat pomocí metod varietního učení

20. Který z uvedených způsobů validace dat je lepší?

- Opakovaná „hold-out“ metoda
- Křížová validace
- Resubstituce



1. Metoda, jež umožňuje spočítat vzdálenost vzdušnou čarou mezi dvěma body.
2. Předpoklad mnoha metod vícerozměrné analýzy dat.
3. Vztah dvou spojených veličin, když s rostoucími hodnotami jedné proměnné narůstají hodnoty druhé proměnné.
4. Metoda, která umožňuje zobrazit pozorované vzdálenosti/nepodobnosti mezi objekty v euklidovském prostoru (zkratka).
5. Tabulkové znázornění vztahu proměnných či objektů.
6. Grafické znázornění vztahu většího počtu spojených proměnných.
7. Metoda vedoucí zpravidla k protáhlým shlukům v datech.
8. Grafické znázornění podobnosti objektů převedením hodnot proměnných na tvary či symboly.
9. Typ analýzy, která umožňuje vytvořit shluky stejného řádu.
10. Metoda, která umožňuje transformaci dat do nových nekorelovaných proměnných, přičemž bude zachováno co nejvíce původní variability (zkratka).
11. Metoda, jež umožňuje zjistit, zda se od sebe liší dvě skupiny objektů, které jsou popsány hodnotami několika proměnných.