Redukcia dimenzie dát
Zadanie projektu

Text projektu:

- odovzdávajú sa 3 súbory:
  - text projektu v PDF (typicky 5-10 strán)
  - Rkovský skript: musí byť spustiteľný a musia sa z neho dať získať všetky výsledky v texte
  - súbor s dátami
- odovzdávanie: uploadnutím do Assignmentu v MS Teams, najneskôr deň pred skúškou (t.j. do 23:55 dňa predchádzajúceho skúške)
- (namiesto Rka teoreticky môže byť aj Python, ale musíte v takom prípade veľmi dobre vedieť vysvetliť, čo ste v Pythone spravili a čo jednotlivé funkcie robia)

Prezentácia:

- Projekt prezentujete iba mne, ako to býva na ústnej skúške. V princípe sa porozprávame o tom, čo ste spravili.
- Neučte sa svoje rozprávanie naspamäť a určite nevyrábajte prezentáciu (pdf či powerpoint).

Obsah:

- Zvoľte si dáta a použite metódy, ktoré preberáme na tomto predmete, na to, aby ste sa dozvedeli **niečo zaujímavého** o zvolených dátach alebo na základe zvolených dát. Napríklad:
  - zaujímavé vykreslenie v menejrozmere
  - efektívne použitie (redukovaných dát) v nejakých iných modeloch (napr. ako vstupy do niektorých metód z AZKD)
  - reprezentovanie výstupov nejakých iných modelov (napr. niektorých metód z AZKD)
  - získanie čo najlepšieho predikčného modelu
- Použite aspoň jednu metódu z časti **A** (str. 3) a aspoň jednu metódu z časti **B** alebo **C**. Keďže "aspoň", tak môžete použiť aj viac metód. Ak z nejakého dôvodu chcete aplikovať inú množinu metód, ozvite sa mi.

Poznámky:

- Projekt by mal mať "slušnú" formu, čiže typicky by mal zahŕňať napríklad popis dát, prvotný pohľad na dáta (napr. scatterplot(y), histogram(y), priemery, kovariancie a korelácie...), popis toho, čo ste robili, výstupov, interpretácií, a kritické zhodnotenie výsledkov.
- Projekt by mal obsahovať aj niečo nad rámec mechanického aplikovania kódov z cvičení: napr. porovnanie rôznych metód, užitočné zdôvodnenie (ne)vhodnosti nejakej metódy na dané dáta, skúšanie rôznych nastavení danej metódy, praktické použitie metód (na vykreslenie výstupov iných metód, na získanie dobrého predikčného modelu...), dôkladne analyzovaná interpretácia/aplikovateľnosť výsledkov, hľadanie najlepšieho vykresľovania výstupu.
- Nie iba vypísať (vykresliť) výsledky z Rka, ale hlavne slovne popisovať a interpretovať.
- Vaše analýzy by mali mať jasnú motiváciu; v texte treba zahrnúť, prečo jednotlivé kroky robíte a čo ste nimi dosiahli.
- Všeobecnú teóriu o metódach nepopisujte.
- Textu by mal rozumieť aj človek, čo nerobí s Rkom.

Dáta:

- reálne, nie vymyslené, nie vygenerované
- môžete ich aj sami získať (napr. dotazník)
- nemôžete používať tie, ktoré sme analyzovali na prednáškach
- môžete používať dáta z vašich záverečných prác alebo z vašich projektov, ktoré ste spravili/robíte na iných predmetoch
- dátam musíte rozumieť (napr. ak vôbec nerozumiete časticovej fyzike, dáta z urýchľovačov častíc nie sú vhodné)

Hodnotí sa:

- vhodnosť a správnosť použitých metód
- správnosť a zrozumiteľnosť interpretácie
- kvalita textu a obrázkov
- kvalita prezentácie
- **originálnosť, zaujímavosť** (dát, analýz) a celkový dojem – čo zaujímavého sa človek z projektu dozvie

Časť A: Základné metódy extrakcie premenných

- analýza hlavných komponentov
- mnohorozmerné škálovanie

Časť B: Zložitejšie metódy extrakcie premenných

- faktorová analýza
- projekčné sledovanie
- nelineárna analýza hlavných komponentov
- Isomap
- t-SNE
- autoencoders
- self-organizing maps

Časť C: Metódy selekcie premenných

- kombinatorické metódy selekcie
- lasso
- hrebeňová regresia