

Redukcia dimenzie dát

Informácie o skúške

(snáď) finálna verzia 14.5.2025

Termíny a miestnosť:

- sú uvedené v AISe
- prihlasovať/odhlasovať sa dá vždy do 8:50 predchádzajúceho dňa

Harmonogram:

č. študenta	príprava	odpoveď
1	8:50-9:30	9:30-10:10
2	9:30-10:10	10:10-10:50
3	10:10-10:50	10:50-11:30
prestávka		
4	12:20-13:00	13:00-13:40
5	13:00-13:40	13:40-14:20
6	13:40-14:20	14:20-15:00
7	14:20-15:00	15:00-15:40
8	15:00-15:40	15:40-16:20

- Môže sa stať, že počas dňa vzniknú časové sklzy.
- Medzi prihlásenými študentami si, prosím, dohodnite poradie. Ak sa nedohodnete, budem predpokladať poradie podľa abecedy.
- Zoznam prihlásených študentov vždy všetkým prihláseným pošlem mailom po tom, ako sa uzavrie prihlasovanie (teda doobeda v deň pred skúškou).

Odpoveď pozostáva z:

1. Vylosovanie dvoch otázok a príprava na odpoveď (40 minút),
 - Typicky počas odpovede predchádzajúceho študenta.
 - Príprava prebieha v tej istej miestnosti ako skúška.
2. Samotná odpoveď (40 minút)
 - (a) Prezentácia projektu (viď http://www.iam.fmph.uniba.sk/ospm/Rosa/RDD/RDD_projekt.pdf)
 - (b) Odpoveď na vylosované otázky.

Poznámky:

- Pri každej metóde je potrebné vedieť aj aký je jej cieľ/použitie a aké sú jej hlavné výhody (prípadne nevýhody, ak boli spomínané).

- Budem sa pýtať aj doplňujúce otázky.

Otázky:

1. Metóda hlavných komponentov: geometria, optimalizačná úloha a jej riešenie (vrátane dôkazu).
2. Metóda hlavných komponentov: voľba počtu komponentov, teoretické hlavné komponenty, interpretácia vl. vektorov, PCA a kolmé projekcie.
3. Klasické mnohorozmerné škálovanie (netreba vedieť Youngovu-Hausholderovu vetu).
4. Nemetrické mnohorozmerné škálovanie.
5. Jadrová PCA (stačí hlavná myšlienka jadrového triku, nie jeho detailné odvodenie).
6. Isomap.
7. t-SNE (netreba formálnu definíciu perplexity).
8. Faktorová analýza (z varimax/kvartimax stačí jedno).
9. Autoencoders.
10. Self-organizing maps.
11. Kvalita predikcie, bias-variance tradeoff. Trénovacia a testovacia vzorka.
12. Kombinatorická selekcia premenných: validácia a krosvalidácia, informačné kritériá, spôsoby selekcie.
13. Hrebeňová regresia.
14. Lasso; Principal component regression.
15. Výber podvzorky: bez ohľadu na model, pre súčin matíc a pre MNŠ odhad.