

## Úkol B – Hledání podskupin v datech – Point data

### Cíl

Cílem druhého úkolu je seznámit se s praktickým použitím shlukovacích metod. Shlukovací metody lze využít k odhalení vnitřní struktury dat. Poznání vnitřní struktury dat může pomoci k lepšímu porozumění datům a jevům, které zachycují, stejně tak ke konstrukci lepších klasifikátorů.

### Data

Point data byla sesbírána v průběhu několika let na více než 2000 pacientech podstupujících náročné lékařské zákroky. Cílem studie bylo posoudit výhody a nevýhody přesné kontroly glykémie pacienta (koncentrace glukózy v krvi) během operace. Velmi nízká koncentrace glukózy (hypoglykémie) stejně jako velmi vysoká koncentrace glukózy (hyperglykémie) mohou být pro pacienty velmi nebezpečné. V rámci studie byly použity dva protokoly při provádění operace, v prvním případě byla hladina glukózy velmi přesně regulována, zatímco v druhém případě nebyla hladina glukózy kontrolována, ale pouze měřena. Z důvodu posouzení výhod byly pacienti sledováni při pobytu na anesteziologicko - resuscitačním oddělení a veškeré pooperační komplikace byly zaznamenány.

### Hledání podskupin v datech

Pokuste se v datech nalézt vnitřní strukturu (podskupiny) na základě hodnot příznaků. Zvolte vhodnou metodu. Zamyslete se nad tím, zda je vhodné data normalizovat. Výsledky vizualizujte a pokuste se je interpretovat, např. pomocí typických (průměrných) reprezentantů. Prozkoumejte, které z příznaků se nejvíce podílí na rozdělení instancí do jednotlivých shluků. Uvažte jestli je možné dát nalezenou strukturu do souvislosti s výskytem pooperačních komplikací.

### Požadované kroky analýzy [10 bodů]

- upravte data pro shlukování [1 b]
- zvolte vhodnou metodu shlukování [1 b]
- zvolte vhodné parametry shlukovací metody [1 b]
- prezentujte výsledky shlukování pomocí typických reprezentantů shluků [3 b]
- zjistěte v jakých příznacích a jak se jednotlivé shluky liší [2 b]
- porovnejte výsledky shlukování s výskytem pooperačních komplikací [2 b]

Výsledky upravte do formy zprávy, která bude obsahovat stručný **úvod**, popis metod, které jste použili, v sekci **metody**, výsledky jejich aplikace na data v sekci **výsledky** a závěry, které jste zjistili interpretací výsledků v sekci **závěr**. Maximální délka zprávy je 3 stránky. Zprávu ve formátu pdf odevzdejte pomocí UploadSystemu.