

Úkol B – Hledání podskupin v datech – Arrhythmia 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

Arrhythmia data byla vytvořena pro účely klasifikace normálního průběhu EKG a 15 arytmií. Informace o pacientech spolu s průběhy 12-ti svodového EKG byly popsány pomocí 279 příznaků. Mezi nimi je prvních 15, které představují obecné pacientské charakteristiky, a jejich popis je tabulce níže. Zbytek jsou parametry z jednotlivých EKG svodů popsané pomocí délek segmentů a jejich amplitud, a pak specifické jevy typické pro arytmiie.

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 klasifikací srdečních arytmií.

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 klasifikací srdečních arytmií [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.