

## Úkol B – Hledání podskupin v datech – Forest fires ZS 2016

### 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

Lesní požár je událost spojená s velkými riziky pro lesní ekonomii a současně ohrožující lidi, i když z hlediska ekologie a životního prostředí se jedná o přirozenou formu obnovy lesních společenství. Zvládnutelnost lesního požáru a jeho následky jsou ovlivněny počasím v době požáru. Data z portugalské studie nabízí zajímavé informace o vztazích mezi jednotlivými parametry počasí a velikostí lesních požárů.

### 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 pozorování do jednotlivých shluků. Uvažte, jestli je možné dát nalezenou strukturu do souvislosti s plochou spáleniště.

### 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 [3 b]
- porovnejte výsledky shlukování s plochou spáleniště [1 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.

<b>X</b>	odvozená x-ová souřadnice (od 1 do 9)
<b>Y</b>	odvozená y-ová souřadnice (od 1 do 9)
<b>month</b>	měsíc v roce (leden až prosinec)
<b>day</b>	den v týdnu (pondělí až neděle)
<b>FFMC</b>	FFMC kód
<b>DMC</b>	DMC kód
<b>DC</b>	DC kód
<b>ISI</b>	ISI index
<b>temp</b>	venkovní teplota (ve °C)
<b>RH</b>	venkovní relativní vlhkost (v %)
<b>wind</b>	venkovní rychlost větru (v km/h)
<b>rain</b>	venovní déšť (v mm/m <sup>2</sup> )
<b>area</b>	celková plocha spáleniště (v ha)

Tabulka 1: Popis proměnných z studie lesních požárů v severovýchodním Portugalsku