

MAD - pracovní list 4

Analýza hlavních komponent

1. část - Společná práce

Analýza hlavních komponent je velmi důležitá metoda při práci s vícerozměrnými daty. Je totiž celkem přirozenou otázkou, zda naše pozorování, resp. měření neobsahují složky, které by šly vyjádřit pomocí ostatních sledovaných veličin a jejichž měření bychom již tedy příště mohli vynechat, popřípadě je zanedbat. Analýza hlavních komponent tak může vést k redukci dimenze našich dat. Její provedení je založené na znalosti některých pojmů z lineární algebry, kterým byl věnován první týden.

2. část - Samostatná práce

1. vizualizujte si graf pro 2. a 3. hlavní komponentu dat olympic, dále pro 3. a 4. komponentu
2. Načtěte si opět data z minulého týdne a proveďte analýzu hlavních komponent pro tato data.
3. Načtěte si data iris
 - budete potřebovat knihovnu **library(ggfortify)**
 - zjistěte si, čeho se data týkají
 - proveďte základní průzkum a vizualizaci dat Iris
 - proveďte analýzu hlavních komponent
 - pomocí analýzy hlavních komponent se pokuste určit optimální počet shluků v datech a proveďte k-means clustering pro tento počet shluků
 - zkuste volit i jiný počet shluků a nakreslete clustergramy