

Podstawy Przetwarzania Danych

Laboratoria 3: Redukcja wymiarowości

dr inż. Marcin Luckner
mluckner@mini.pw.edu.pl

Wydział Matematyki i Nauk Informacyjnych

Wersja 1.1
5 marca 2021

Projekt „NERW 2 PW. Nauka – Edukacja – Rozwój – Współpraca” współfinansowany jest ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego.

Zadanie 10 pn. „Modyfikacja programów studiów na kierunkach prowadzonych przez Wydział Matematyki i Nauk Informatycznych”, realizowane w ramach projektu „NERW 2 PW. Nauka – Edukacja – Rozwój – Współpraca”, współfinansowanego jest ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego.

Wyliczanie składowych głównych

```
1 waw.pca <- prcomp(x, center = TRUE, scale. = TRUE)
```

- Metoda `prcomp` wylicza składowe główne dla danych `x`.
- Dane można standaryzować używając zmiennych `center` i `scale`.

Wyświetlanie PCA

```
1 print(waw.pca)
```

- Metoda `print` wyświetla współczynniki liniowej kombinacji zmiennych tworzących składowe.

Analiza PCA

```
1 summary(waw.pca)
```

- Metoda `summary` dostarcza informacji o zmienności czynników i procencie tłumaczonych danych.

Biblioteka ggbiplot

```
1 require(ggbiplot)
2 ggbiplot(waw.pca)
3 ggscreeplot(waw.pca)
```

- Biblioteka ggbiplot pozwala tworzyć wykresy dotyczące PCA.
 - Metoda ggbiplot pokazuje jak zmienne tworzą pierwsze dwie składowe
 - Metoda ggscreeplot tworzy wykres osypiskowy.

Zadanie 1

- Dla danych z pliku wawData.csv wykonaj wybór liczby składowych postępując się:
 1. Kryterium własności własnej,
 2. Kryterium części wyjaśnionej wariancji (90%),
 3. Kryterium wykresu osypiskowego.

Kryterium minimalnego zasobu zmienności wspólnej

- W celu zastosowania kryterium minimalnego zasobu zmienności wspólnej należy
 1. Policzyc kwadraty współczynników zwracanych przez metodę print
 2. Policzyc skumulowane wartości kwadratów dla kolejnych składowych.

Zadanie 2

- Sprawdź, czy kryterium minimalnego zasobu zmienności wspólnej jest zgodne z wartościami z poprzednich kryteriów.
- Jaką liczbę składowych powinniśmy wykorzystać i dlaczego?

Analiza czynnikowa

```
1 waw.fit <- factanal(x, factors=2, rotation="none",  
  scores="Bartlett")
```

- Metoda `factanal` tworzy model dla danych `x`.
- Model jest tworzony dla liczby czynników określonych parametrem `factors`.
- Poprawność modelu jest badana przy użyciu metody określonej parametrem `scores`.
- Można ustalić rotację czynników parametrem `rotation`

Analiza modelu

```
1 print(waw.fit , digits=2, cutoff=.3, sort=TRUE)
```

- Metoda `print` prezentuje utworzony model.
- Parametr `digits` określa dokładność prezentowania danych.
- Parametr `cutoff` określa wartość poniżej której czynniki są zerowane.
- Ustawienie parametru `sort` na `TRUE` powoduje uporządkowanie czynników względem istotności.

Rotacja

```
1 fitRot <- factanal(x, factors=2, rotation="varimax",  
  scores="Bartlett")
```

- Ustalenie wartości parametru `rotation` na `varimax` wprowadza rotację czynników.

Ocena modelu

- Sposób oceny modelu zależy od parametru α .
- Ocena zwraca liczbę stopni swobody określającą liczbę niezależnych zmiennych losowych, które wpływają na wynik.
- Jeżeli liczba stopni swobody jest większa niż zero model wylicza wartość p .
- Ocenę czynników przeprowadzamy analizując sumę kwadratów ładunków (SS loading). Jeżeli przekracza on 1 to czynnik jest użyteczny.

Zadanie 3

- Wykonaj analizę czynnikową używając testu Bartletta.
- Jaka jest maksymalna liczba czynników, które mogą być użyte?
- Które zmienne nie są reprezentowane przez model dla parametru $\text{cutoff}=0.3$?

Zadanie 4

- Wykonaj ponownie analizę czynnikową używając tych samych parametrów i rotacji varimax.
- W jaki sposób zmieniły się parametry modelu?