



Aalto-yliopisto
Perustieteiden
korkeakoulu

Poikkeavien havaintojen tunnistaminen pääkomponenttianalyysin menetelmin (aihe-esittely)

Kalle Alahuusua

01.09.2018

Ohjaaja: *TkT Lauri Viitasaari*

Valvoja: *Apulaisprofessori Pauliina Ilmonen*

Työn saa tallentaa ja julkistaa Aalto-yliopiston avoimilla verkkosivuilla. Muilta osin kaikki oikeudet pidätetään.

Pääkomponenttianalyysi

- Laaditaan joukko keskenään korreloimattomia uusia muuttujia lineaarisesti yhdistelemällä alkuperäisiä muuttujia
- Kuvaa havaintojoukon varianssi-kovarianssirakennetta
- Ensimmäinen pääkomponentti maksimoi projektoidun datan varianssin ja minimoi residuaalit datapisteiden ja projektoiden välillä
- Pääkomponentit kovarianssimatriisin ominaisvektoreita
- Herkkä poikkeaville havainnoille

Vakaat estimaattorit

- Pyrkivät kuvaamaan otantajakauman keskeisiä piirteitä huolimatta poikkeavien havaintojen läsnäolosta
- Normaalisuusoletuksesta luovutaan
- Tasapainottelua tehokkuuden ja vakauden välillä
- Esim. keskiarvo vs. Mediaani

Vakaa pääkomponenttianalyysi

Kaksi koulukuntaa:

1. Korvataan klassinen kovarianssimatriisi vakaalla estimaatilla
 2. Maksimoidaan vakaata hajontaestimaattia, ja haetaan pääkomponenttien suunnat iteratiivisesti
- Sama menetelmä voi hyödyntää kumpaakin periaatetta

Tavoite

- Tunnistetaan poikkeavat havainnot epäsymmetrisestä havaintojoukosta tuotantotekniikan alalta

Ideaalitilanne:

- Poikkeavat havainnot sijaitsevat poikkeuksellisen kaukana havaintojoukon vakaasta keskipisteestä sekä vakaiden pääkomponenttien virittämästä tasosta

Hubert, M., P. J. Rousseeuw, and K. V. Branden (2005).
"ROBPCA: A New Approach to Robust Principal Component Analysis"

Menetelmät

ROBPCA

- Yhdistelee kahta vakaan pääkomponenttianalyysin koulukuntaa
- Menetelmästä kehitetty muunnelma epäsymmetrisen datan analyysiin
- Implementoitu kattavasti (R, MATLAB)

Työssä käytämme:

LIBRA: the MATLAB Library for Robust Analysis
(Verboven ja Hubert 2010)

Aikataulu

- Aiheen esittely 01.09.2018
- Valmiin työn esittely seminaarissa 03.10.2018

Viitteet

Hubert, M., P. J. Rousseeuw ja K. V. Branden (2005),
“ROBPCA: A New Approach to Robust Principal Component Analysis”,
Technometrics 47.1, sivut 64–79.