



Aalto-yliopisto
Perustieteiden
korkeakoulu

Spatiaalisiin merkkeihin ja järjestyslukuihin perustuva moniulotteinen regressioanalyysi R-ohjelmistoa käyttäen (aihe-esittely)

Niko Lietzén

3.3.2014

Ohjaaja/valvoja: *apulaisprof. Pauliina Ilmonen*

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

Tausta

- L_2 -normiin perustuvat perinteiset regressiomenetelmät ovat optimaalisia, kun residuaalit noudattavat moniulotteista normaalijakaumaa
- Ne ovat kuitenkin herkkiä poikkeaville havainnoille
- Eivät myöskään optimaalisia paksuhäntäiselle datalle

Tausta

- L_1 -normiin perustuvat menetelmät ovat robustimpeja, eli eivät niin herkkiä poikkeaville havainnoille

Tavoitteet

- Vertaillaan L_2 -regressiomenetelmää ja robustimpaa vaihtoehtoa
- Pääpaino spatiaalisiin merkkeihin ja järjestyslukuihin perustuvilla menetelmillä
- Sovelletaan menetelmiä simuloituun moniulotteiseen elliptiseen- ja normaalijakaumaan, joihin on lisätty poikkeavia arvoja

Menetelmät

- Lineaarinen regressiomalli

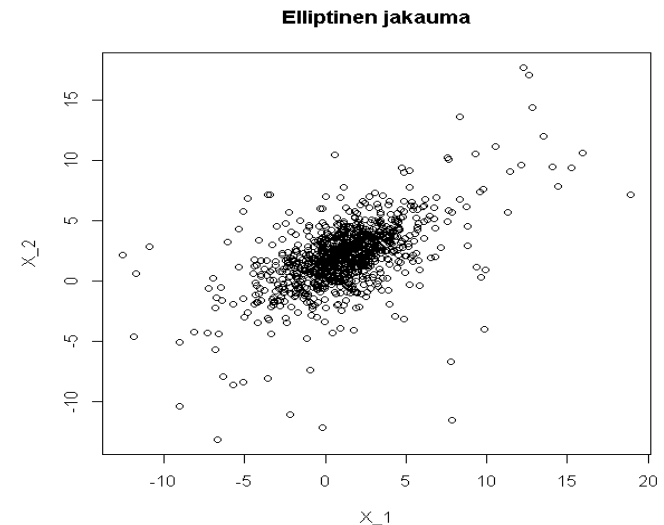
$$Y = X\beta + \varepsilon$$

- Estimoidaan β pistemääräfunktion T avulla

$$T(\hat{\beta})'X = 0$$

- L_2 regressio

$$(Y - X\hat{\beta})'X = 0$$



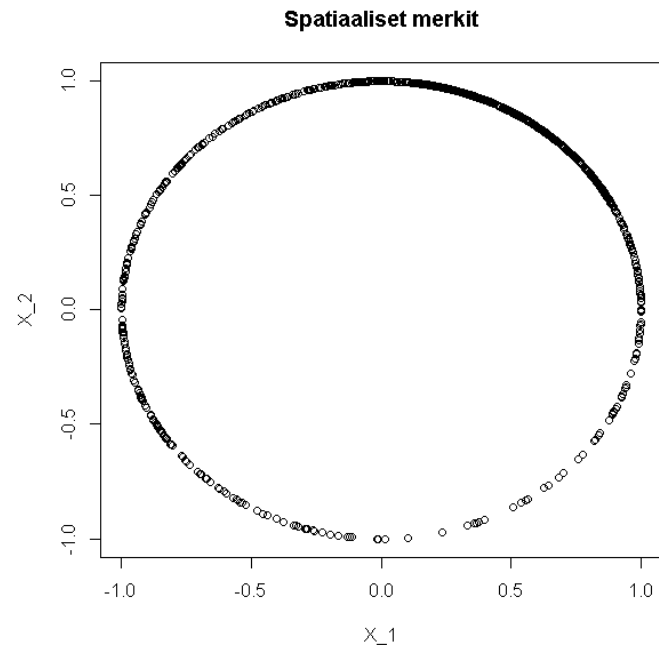
Menetelmät

- Spatiaalinen merkki
 - Kertoo datapisteen suunnan suhteessa origoon
 - Vektori, jonka pituus on 1
- Määritelmä

$$U(y) = \begin{cases} |y|^{-1}y, & y \neq 0 \\ 0, & y = 0 \end{cases}$$

- Pistemääräfunktio

$$U(\hat{\beta})'X = 0$$



Menetelmät

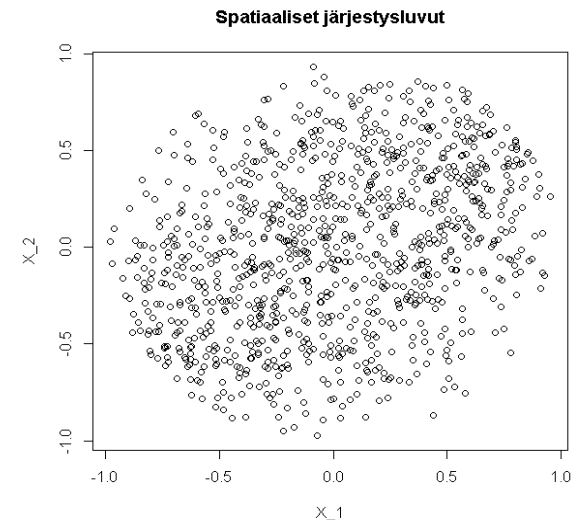
- Spatiaalinen järjestysluku
 - Vektori, jonka suunta kertoo missä datapiste likimain sijaitsee suhteessa dataparven keskipisteeseen
 - Pituus kertoo kuinka kaukana datapiste likimain sijaitsee dataparven keskipisteestä

- Määritelmä

$$R(y) = AVE\{U(y - y_i)\}$$

- Pistemääräfunktio

$$R(\hat{\beta})'X = 0$$



Työkalut

- Tärkein työkalu R-ohjelmisto
- Käytetään pakettia MNM (Multivariate Nonparametric Methods)
 - Valmis paketti spatiaalisten merkkien ja järjestyslukujen hyödyntämiseen

Aikataulu

- Aiheen esittely 3.3.2014
- Aihe 12/2013
- Kirjoittaminen 1/2014
- Valmiin työn esittely 28.4.2014
- Työn laajentaminen Kesä 2014?

Tietolähteet

- Journal of Statistical Software: Multivariate L_1 Methods: The package MNM
 - Nordhausen & Oja, heinäkuu 2011
 - MNM-paketin esimerkit ja ohjeet
 - <http://cran.r-project.org/web/packages/MNM/MNM.pdf>
 - Multivariate Analysis
 - Mardia, Kent & Bibby, 2003
 - Multivariate Nonparametric Methods with R
 - Oja, 2010
-