Ei-parametrinen regressio
Harjoitus 4
15.4.2008

1. Osoita, että regressiomallissa, kun käytetään pienimmän neliösumman menetelmää
$$\sum Var(\hat{\beta}_j) > \sigma^2 \rho_{min}^{-1},$$
missä ρ_{min} on matriisin $\mathbf{X}'\mathbf{X}$ pienin ominaisarvo.

2. Osoita, että minimoitavan funktion
$$PLS(\beta) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \lambda\boldsymbol{\beta}'\boldsymbol{\beta},$$
missä $\lambda > 0$, ratkaisuna saadaan ns. harja-estimaattori
$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}.$$

3. Jos harja-estimaattorin tapauksessa määritellään
$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y},$$
missä $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}'$, niin osoita, että
$$df = tr(\mathbf{H}) = \sum \frac{\rho_j}{\rho_j + \lambda},$$
missä ρ_j on matriisin $\mathbf{X}'\mathbf{X}$ j. ominaisarvo. Kokeile ennustamista (piirrä kuvio) eri vapausasteiden df arvoilla valitsemassasi aineistossa.

4. Estimoi aineistoon *faithful* paloittain lineaarinen malli siten, että kertoimia on tasoitettu sekamallin avulla. Vertaa graafisesti saatua sovitetta tasoittamattomaan.

5. Tarkastellaan minimoitavaa lauseketta
$$PLS = (\mathbf{y} - \mathbf{g})'(\mathbf{y} - \mathbf{g}) + \lambda\mathbf{g}'\mathbf{Q}\mathbf{Q}'\mathbf{g},$$
Osoita, että ratkaisu voidaan kirjoittaa muodossa
$$\hat{\mathbf{g}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}} = (\mathbf{I} + \lambda\mathbf{Q}\mathbf{Q}')^{-1}\mathbf{y},$$
missä
$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

ja

$$\hat{\mathbf{u}} = (\mathbf{I} + \lambda \mathbf{Q}\mathbf{Q}')^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}),$$

kun

$$\mathbf{Q} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ -2 & 1 & \ddots & 0 & 0 \\ 1 & -2 & 1 & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & & & 1 \\ & & & & -2 \\ 0 & 0 & \cdots & & 1 \end{pmatrix}.$$