Topics in Statistical Sciences

Mikkel Findinge

Indhold

Eksamensopgaver 1

Sørens del - se `R`-kode.

Eksamensopgaver 2

1. Vis, at ved at udvide responsen $\mathbf{y}$ og design matricen $\mathbf{X}$ på tilpas vis kan ridge regressions estimatoren opnås fra OLS udtrykket $(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}\widetilde{\mathbf{y}}$, hvor $\widetilde{\mathbf{y}}$ og $\widetilde{\mathbf{X}}$ er udvidelserne.

Ridge regressions estimatoren er givet ved

$$\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_2^2.$$

Vi kan lave små omskrivninger

$$\begin{aligned}\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_2^2 &= \min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \|\sqrt{\lambda}\boldsymbol{\beta}\|_2^2 \\ &= \min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \|0 - \sqrt{\lambda}I\boldsymbol{\beta}\|_2^2\end{aligned}$$

Sætter vi $\widetilde{\mathbf{y}} = [\mathbf{y} \quad 0]^\top$ og $\widetilde{\mathbf{X}} = [\mathbf{X} \quad \sqrt{\lambda}I]^\top$ kan vi opskrive:

$$\min_{\boldsymbol{\beta}} \|\widetilde{\mathbf{y}} - \widetilde{\mathbf{X}}\boldsymbol{\beta}\|_2^2.$$

Nu ved vi altså, hvordan de udvidede vektorer og matricer ser ud for at omskrive ridge regression til ordinary least squares. Vi betragter nu OLS udtrykket:

$$\widehat{\boldsymbol{\beta}} = (\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{y}} = \left([\mathbf{X}^\top \quad I\sqrt{\lambda}]\begin{bmatrix}\mathbf{X} \\ \sqrt{\lambda}I\end{bmatrix}\right)^{-1}[\mathbf{X}^\top \quad \sqrt{\lambda}I]\begin{bmatrix}\mathbf{y} \\ 0\end{bmatrix} = (\mathbf{X}^\top\mathbf{X}+\lambda I)^{-1}\mathbf{X}^\top\mathbf{y}.$$

Hvilket var, hvad vi skulle vise.

2. Vis hvordan $\ell_Q(\beta_0, \boldsymbol{\beta})$ opnås fra

$$\ell(\beta_0, \boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\left[y_i(\beta_0 + \mathbf{X}_i^\top\boldsymbol{\beta}) - \log(1 + \exp\{\beta_0 + \mathbf{X}_i^\top\boldsymbol{\beta}\})\right].$$

Not gonna write this one. Not gonna happen.

3. Udled opdateringsligningerne for β_j baseret på $-\ell_Q + \lambda P_\alpha(\boldsymbol{\beta})$ for $j = 1, \ldots, p$.

Først differentieres $\ell_Q = -\frac{1}{2n}\sum_{i=1}^{n} w_i^*(z_i^* - \beta_0 - x_i^\top\boldsymbol{\beta})^2$ i forhold til β_j. Vi får ved kædereglen, at:

$$\frac{\partial}{\partial\beta_j}\ell_Q = -\frac{1}{2n}\sum_{i=1}^{n} -2x_{ij}w_i^*(z_i^* - \beta_0 - x_i^\top\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} x_{ij}w_i^*(z_i^* - \beta_0 - x_i^\top\boldsymbol{\beta}).$$

Da vi i sidste ende vil isoleres β_j kan vi lige lave en sum udelukkende bestående af led, der indeholder β_j. Vi har:

$$\frac{1}{n}\sum_{i=1}^{n} x_{ij}w_i^*(z_i^*-\beta_0-x_i^\top\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} x_{ij}w_i^*(z_i^*-\beta_0-\sum_{k\neq j}(x_{ik}\beta_j))-\frac{1}{n}\beta_j\sum_{i=1}^{n} x_{ij}^2w_i^* = a_j-\beta_j c_j$$

Vi har desuden, at

$$\partial|\beta_j| = \begin{cases} -1 & \beta_j < 0 \\ [-1,1] & \beta_j = 0 \\ 1 & \beta_j > 0, \end{cases}$$

hvormed

$$\frac{\partial}{\partial\beta_j}P_\alpha(\boldsymbol{\beta}) = \frac{\partial}{\partial\beta_j}\left(\sum_{i=1}^{p}(1-\alpha)\beta_i^2/2+\alpha|\beta_i|\right) = (1-\alpha)\beta_j+\alpha\partial|\beta_j|.$$

Vi samler de forskellige differentialer:

$$\frac{\partial}{\partial\beta_j}-\ell_Q+\lambda P_\alpha(\boldsymbol{\beta}) = -a_j+\beta_j c_j+\lambda((1-a)\beta_j+\alpha\partial|\beta_j|) = \beta_j(c_j+\lambda(1-\alpha))-a_j+\lambda\alpha\partial|\beta_j| = 0.$$

Vi isolerer β_j og får:

$$\widehat{\beta}_j = \frac{a_j-\lambda\alpha\partial|\beta_j|}{c_j+\lambda(1-\alpha)}$$

Måden, c_j er defineret på, gør denne konstant positiv. Dette er $\lambda(1-\alpha)$ også. Dermed er nævneren altid positiv. Det vil sige, at det eneste, der kan ændre fortegnet på β_j er tælleren. Vi har:

$$\widehat{\beta}_j = \frac{a_j-\lambda\alpha\partial|\beta_j|}{c_j+\lambda(1-\alpha)} = \frac{1}{c_j+\lambda(1-\alpha)}\begin{cases} a_j+\lambda\alpha & a_j < -\lambda\alpha \\ 0 & -\lambda\alpha \leq a_j \leq \lambda\alpha \\ a_j-\lambda & a_j > \lambda\alpha. \end{cases}$$

Hvis man har lyst, kan man skrive det som:

$$\widehat{\beta}_j = \frac{\operatorname{sign}(a_j)(|a_j|-\lambda\alpha)_+}{c_j+\lambda(1-\alpha)}.$$

Eksamensopgaver 3

Man kan estimere $\boldsymbol{\beta}$ ved at løse følgende GEEs

$$\psi = \sum_{i=1}^{K} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}^{\top} V_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}))$$

for

$$\Sigma = \phi A_i^{1/2} R(\alpha) A_i^{1/2},$$

hvor A_i er en $n_i \times n_i$ diagonalmatrix med $vV(\mu_{ij})$ på den j'te diagonalindgang og $R(\alpha)$ er korrelationsmatricen.

I opgave 1 betragtes følgende setup

$$V(\mu) = 1, \quad g(\mu) = X\boldsymbol{\beta}, \quad R(\alpha) = I, \quad w_i = 1.$$

Se `R`-kode for resten.

Eksamensopgaver 4

1a. Vis, at $S(\boldsymbol{\beta}) = \|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|_2^2$ kan blive dekomponeret, sådan at

$$S(\boldsymbol{\beta}) = \|Q_1^\top(\mathbf{X}\boldsymbol{\beta} - \mathbf{y})\|_2^2 + \|Q_2^\top \mathbf{y}\|_2^2.$$

Den ortogonale matrix $Q = [Q_1 \quad Q_2]$ er en rotation. Denne vil altså ikke påvirke $S(\boldsymbol{\beta})$, som er en længde. Vi har:

$$S(\boldsymbol{\beta}) = \|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|_2^2 = \|Q^\top(\mathbf{X}\boldsymbol{\beta} - \mathbf{y})\|_2^2 = (\mathbf{X}\boldsymbol{\beta} - \mathbf{y})^\top QQ^\top(\mathbf{X}\boldsymbol{\beta} - \mathbf{y}).$$

Da $Q = [Q_1 \quad Q_2]$ er $QQ^\top = [Q_1 \quad Q_2][Q_1 \quad Q_2]^\top = Q_1Q_1^\top + Q_2Q_2^\top$. Men så kan ovenstående omskrives til

$$(\mathbf{X}\boldsymbol{\beta} - \mathbf{y})^\top QQ^\top(\mathbf{X}\boldsymbol{\beta} - \mathbf{y}) = (\mathbf{X}\boldsymbol{\beta} - \mathbf{y})^\top(Q_1Q_1^\top + Q_2Q_2^\top)(\mathbf{X}\boldsymbol{\beta} - \mathbf{y}).$$

Vi ganger ind på hver af ledene i parentesen, således vi får:

$$(\mathbf{X}\boldsymbol{\beta} - \mathbf{y})^\top Q_1Q_1^\top(\mathbf{X}\boldsymbol{\beta} - \mathbf{y}) + (\mathbf{X}\boldsymbol{\beta} - \mathbf{y})^\top Q_2Q_2^\top(\mathbf{X}\boldsymbol{\beta} - \mathbf{y}) = \|Q_1^\top(\mathbf{X}\boldsymbol{\beta} - \mathbf{y})\|_2^2 + \|Q_2^\top(\mathbf{X}\boldsymbol{\beta} - \mathbf{y})\|_2^2.$$

Husk, at vi arbejder med en QR-faktorisering af X, hvor $R = [R_1 \quad 0]^\top$, hvorfor $X = QR = [Q_1 \quad Q_2][R_1 \quad 0]^\top = Q_1R_1$. Vi omskriver således:

$$\|Q_1^\top(\mathbf{X}\boldsymbol{\beta} - \mathbf{y})\|_2^2 + \|Q_2^\top(\mathbf{X}\boldsymbol{\beta} - \mathbf{y})\|_2^2 = \|Q_1^\top(\mathbf{X}\boldsymbol{\beta} - \mathbf{y})\|_2^2 + \|Q_2^\top(Q_1R_1\boldsymbol{\beta} - \mathbf{y})\|_2^2.$$

Men $Q = [Q_1 \quad Q_2]$ er en ortogonal matrix, hvorfor søjlevektorerne er indbyrdes ortogonale. Dermed er $Q_2^\top Q_1 = 0$. Vi får ved at gange ind i sidste parentes:

$$\|Q_1^\top(\mathbf{X}\boldsymbol{\beta} - \mathbf{y})\|_2^2 + \|Q_2^\top Q_1R_1\boldsymbol{\beta} - Q_2^\top\mathbf{y}\|_2^2 = \|Q_1^\top(\mathbf{X}\boldsymbol{\beta} - \mathbf{y})\|_2^2 + \|Q_2^\top\mathbf{y}\|_2^2.$$

1b. For $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$, vis at $S(\hat{\boldsymbol{\beta}}) = \|Q_2^\top\mathbf{y}\|_2^2$.

Vi indsætter blot estimatet for $\boldsymbol{\beta}$ i dekomponeringen fra før. Vi skal blot vise, at

$$\|Q_1^\top(\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} - \mathbf{y})\| = 0.$$

Vi omskriver og får:

$$(\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} - \mathbf{y})^\top Q_1Q_1^\top(\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} - \mathbf{y}).$$

Husk, at $\mathbf{X} = Q_1R_1$, hvor R_1 er en $p \times p$-matrix med fuld rank (ergo invertibel), dermed kan vi skrive $\mathbf{X}R_1^{-1} = Q_1$, hvorfor vi får:

$$(\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} - \mathbf{y})^\top XR_1^{-1}R_1^{-1\top}\mathbf{X}^\top(\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} - \mathbf{y}).$$

Ganger vi $R_1^{-1\top}\mathbf{X}^\top$ ind i parentesen yderst til højre fås:

$$R_1^{-1\top}\mathbf{X}^\top(\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} - \mathbf{y}) = R_1^{-1\top}\mathbf{X}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} - R_1^{-1\top}\mathbf{X}^\top\mathbf{y}$$

Bemærk, at $\mathbf{X}^\top\mathbf{X}$ er ganget på $(\mathbf{X}^\top\mathbf{X})^{-1}$, altså går disse ud med hinanden. Vi har:

$$R_1^{-1\top}\mathbf{X}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} - R_1^{-1\top}\mathbf{X}^\top\mathbf{y} = R_1^{-1\top}\mathbf{X}^\top\mathbf{y} - R_1^{-1\top}\mathbf{X}^\top\mathbf{y} = 0.$$

Dermed har vi vist, at

$$\|Q_1(\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y})\| = 0.$$

2a. Lad $y_i = f(x_i, (\alpha, \beta))$, hvor f er en rationel funktion. Angiv en procedure til at opnå gode startværdier for α og β baseret på data (x, y).

Vi betragter

$$y_i = \frac{\alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + \ldots + \alpha_p x_i^p}{\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_q x_i^q},$$

hvor $\beta_0 \equiv 1$ for at undgå overparametrisering. Vi ganger med nævneren på begge sider og får:

$$y_i(1 + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_q x_i^q) = \alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + \ldots + \alpha_p x_i^p,$$

Gang y_i ind i parentesen og få:

$$y_i + \beta_1 x_i y_i + \beta_2 x_i^2 y_i + \ldots + \beta_q x_i^q y_i = \alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + \ldots + \alpha_p x_i^p,$$

Vi isolerer y_i og får:

$$y_i = \alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + \ldots + \alpha_p x_i^p - \beta_1 x_i y_i - \beta_2 x_i^2 y_i - \ldots - \beta_q x_i^q y_i.$$

For

$$\alpha = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_q \end{bmatrix},$$

kan vi opstille ligningen på formen:

$$y_i = \begin{bmatrix} 1 & x_i & x_i^2 & \ldots & x_i^p & x_i y & x_i^2 y & \ldots & x_i^q y \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}.$$

Vi kan altså for $N = p + q + 1$ par af (x, y)-punkter estimere parametrene ved at løse følgende sæt af ligningssystemer:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^p & x_1 y & x_1^2 y & \ldots & x_1^q y \\ 1 & x_2 & x_2^2 & \ldots & x_2^p & x_2 y & x_2^2 y & \ldots & x_2^q y \\ & & & & \vdots & & & & \\ 1 & x_N & x_N^2 & \ldots & x_N^p & x_N y & x_N^2 y & \ldots & x_N^q y \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}.$$