Solutions to extra exercises for STK2100

Geir Storvik

Sprint 2021

Exercise 1 (Linear regression)

(a) First of all, we have $Y_i \sim N(\boldsymbol{x}_i^T\boldsymbol{\beta}, \sigma^2)$. Since we assume independence, we have that the log-likelihood is

$$\begin{aligned}
l(\boldsymbol{\beta}, \sigma^2) =& \log f(\boldsymbol{y}; \boldsymbol{\beta}, \sigma^2)\\
=& \sum_{i=1}^n \log f(y_i; \boldsymbol{\beta}, \sigma^2)\\
=& \sum_{i=1}^n [-0.5\log(2\pi) - 0.5\log\sigma^2 - 0.5\frac{1}{\sigma^2}(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2]\\
=& -0.5n\log(2\pi) - 0.5n\log\sigma^2 - 0.5\frac{1}{\sigma^2}\sum_{i=1}^n (y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2
\end{aligned}$$

We see that the only term that involves $\boldsymbol{\beta}$ is the last one so that maximizing the (log-)likelihood is equivalent to minimizing

$$\sum_{i=1}^n (y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2 = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$$

(b) We have that

$$(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) = \boldsymbol{Y}^T\boldsymbol{Y} - 2\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{Y} + \boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta}$$

and

$$\frac{\partial}{\partial\boldsymbol{\beta}^T}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) = -2\boldsymbol{X}^T\boldsymbol{Y} + 2\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta}$$

which is equal to zero for

$$\widehat{\boldsymbol{\beta}} = [\boldsymbol{X}^T\boldsymbol{X}]^{-1}\boldsymbol{X}^T\boldsymbol{Y}$$

(c) We have that

$$\frac{\partial}{\partial\sigma^2} l(\boldsymbol{\beta},\sigma^2) = -0.5\frac{n}{\sigma^2} + 0.5\frac{1}{\sigma^4}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2$$

which is equal to zero if

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2$$

resulting in that the maximum likelihood estimate becomes

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}})^2$$

(d) Define $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{Z} + \boldsymbol{b}$. Then

$$Y_j = \sum_k A_{jk}Z_k + b_j E[Y_j] = E[\sum_k A_{jk}Z_k + b_j] = \sum_k A_{jk}E[Z_k] + b_j$$

which combined gives $\mathsf{E}(\boldsymbol{A}\boldsymbol{Z} + \boldsymbol{b}) = \boldsymbol{A}\mathsf{E}(\boldsymbol{Z}) + \boldsymbol{b}$.

Further, defining $\boldsymbol{\Sigma} = \mathsf{V}[\boldsymbol{Z}]$, we have

$$\begin{aligned}
\text{Var}[Y_j] =& \text{Var}[\sum_k A_{jk}Z_k + b_j] = \text{Var}[\sum_k A_{jk}Z_k] \\
=& \sum_k\sum_l \text{Var} A_{jk}A_{jl}\text{Cov}[Z_k, Z_l] = \sum_k\sum_l \text{Var} A_{jk}A_{jl}\Sigma_{k,l} \\
\text{Cov}[Y_j, Y_m] =& \text{Cov}[\sum_k A_{jk}Z_k + b_j, \sum_l A_{ml}Z_l + b_m] \\
=& \text{Cov}[\sum_k A_{jk}Z_k, \sum_l A_{ml}Z_l] \\
=& \sum_k\sum_l A_{jk}A_{ml}\text{Cov}[Z_k, Z_l] = \sum_k\sum_l A_{jk}A_{ml}\Sigma_{k,l}
\end{aligned}$$

which combined gives $\mathsf{V}(\boldsymbol{A}\boldsymbol{Z} + \boldsymbol{b}) = \boldsymbol{A}\mathsf{V}(\boldsymbol{Z})\boldsymbol{A}^T = \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T$.

(e) We know that variances always are positive. Now consider $Z = \boldsymbol{a}^T\boldsymbol{V}$. Then $\mathsf{V}[Z] = \boldsymbol{a}^T\boldsymbol{\Sigma}\boldsymbol{a} \geq 0$ which proves that $\boldsymbol{\Sigma}$ is positive (semi-)definite.

(f) We have that

$$\begin{aligned}
E[\boldsymbol{Y}] =& E[\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}] = \boldsymbol{X}\boldsymbol{\beta} \\
E[\widehat{\boldsymbol{\beta}}] =& E[[\boldsymbol{X}^T\boldsymbol{X}]^{-1}\boldsymbol{X}^T\boldsymbol{Y}] = [\boldsymbol{X}^T\boldsymbol{X}]^{-1}\boldsymbol{X}^T E[\boldsymbol{Y}] \\
=& [\boldsymbol{X}^T\boldsymbol{X}]^{-1}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\beta}
\end{aligned}$$

(*g*) We have

$$\begin{aligned}\mathsf{V}[\widehat{\boldsymbol{\beta}}] =& \mathsf{V}[[\boldsymbol{X}^T\boldsymbol{X}]^{-1}\boldsymbol{X}^T\boldsymbol{Y}]\\ =&[\boldsymbol{X}^T\boldsymbol{X}]^{-1}\boldsymbol{X}^T\mathsf{V}[\boldsymbol{Y}]\boldsymbol{X}[\boldsymbol{X}^T\boldsymbol{X}]^{-1}\\ =&[\boldsymbol{X}^T\boldsymbol{X}]^{-1}\boldsymbol{X}^T\sigma^2\boldsymbol{I}\boldsymbol{X}[\boldsymbol{X}^T\boldsymbol{X}]^{-1}\\ =&\sigma^2[\boldsymbol{X}^T\boldsymbol{X}]^{-1}\boldsymbol{X}^T\boldsymbol{X}[\boldsymbol{X}^T\boldsymbol{X}]^{-1} = \sigma^2[\boldsymbol{X}^T\boldsymbol{X}]^{-1}\end{aligned}$$

Exercise 9 (Multinomial regression)

(*a*) We have for $k > 0$

$$\begin{aligned}\Pr(Y=k|\boldsymbol{x}) =& \frac{\exp(\theta_{k,0}+\sum_{j=1}^p\theta_{k,j}x_j)}{\sum_{l=0}^{K-1}\exp(\theta_{l,0}+\sum_{j=1}^p\theta_{l,j}x_j)}\\ =&\frac{\exp(\theta_{k,0}-\theta_{0,0}+\sum_{j=1}^p(\theta_{k,j}-\theta_{0,j})x_j)}{\sum_{l=0}^{K-1}\exp(\theta_{l,0}-\theta_{0,0}+\sum_{j=1}^p(\theta_{l,j}-\theta_{0,1})x_j)}\\ =&\frac{\exp(\theta_{k,0}-\theta_{0,0}+\sum_{j=1}^p(\theta_{k,j}-\theta_{0,j})x_j)}{1+\sum_{l=1}^{K-1}\exp(\theta_{l,0}-\theta_{0,0}+\sum_{j=1}^p(\theta_{l,j}-\theta_{0,1})x_j)}\\ =&\frac{\exp(\beta_{k,0}+\sum_{j=1}^p\beta_{k,j}x_j)}{1+\sum_{l=1}^{K-1}\exp(\beta_{l,0}+\sum_{j=1}^p\beta_{l,j}x_j)},\end{aligned}$$

for $\beta_{k,j} = \theta_{k,j} - \theta_{0,j}$. The last equation follows by the sum to 1 constraint.

(*b*) We have that

$$\begin{aligned}\Pr(Z_i^k = 1|\boldsymbol{x}_i) =& \Pr(Y_i = k|Y_i = k \text{ or } Y_i = 0, \boldsymbol{x}_i)\\ =&\frac{\Pr(Y_i=k|\boldsymbol{x}_i)}{\Pr(Y_i=k|\boldsymbol{x}_i)+\Pr(Y_i=0|\boldsymbol{x}_i)}\\ =&\frac{\frac{\exp(\beta_{k,0}+\sum_{j=1}^p\beta_{k,j}x_j)}{1+\sum_{l=1}^{K-1}\exp(\beta_{l,0}+\sum_{j=1}^p\beta_{l,j}x_j)}}{\frac{\exp(\beta_{k,0}+\sum_{j=1}^p\beta_{k,j}x_j)}{1+\sum_{l=1}^{K-1}\exp(\beta_{l,0}+\sum_{j=1}^p\beta_{l,j}x_j)}+\frac{1}{1+\sum_{l=1}^{K-1}\exp(\beta_{l,0}+\sum_{j=1}^p\beta_{l,j}x_j)}}\\ =&\frac{\exp(\beta_{k,0}+\sum_{j=1}^p\beta_{k,j}x_j)}{\exp(\beta_{k,0}+\sum_{j=1}^p\beta_{k,j}x_j)+1}\end{aligned}$$

which means that if we only are considering these specific data, we have a standard logistic regression model and the β's can be estimated by standard procedures.

(*c*) See `Extra_9.R`

(*d*) See `Extra_9.R`