Technische Universität München
Zentrum Mathematik
Lehrstuhl für Mathematische Statistik
Prof. Claudia Czado, Ph.D. (cczado@ma.tum.de)
Ariane Hanebeck (ariane.hanebeck@tum.de)

WS 2022/2023

# Exercise Sheet 2
# Generalized Linear Models

**Discussion of the tutorial exercises on November 31 and November 3, 2022**

**Problem 1 (*)** Prove that the coefficient of determination $R^2$ of a linear model is non-decreasing in the number of covariates.

**Problem 2 (*)** We resume our analysis of the USCrime data from Problem 2 on Sheet 1. For this exercise, we use all the covariates `Age, S, Ed, Ex0, Ex1, M, N, U1, U2, W, X`, including the transformed covariates `poly(LF,2)` and `poly(NW,3)`.

a) In Problem 2 on Sheet 1, we saw that not all covariates are significant. Select the best linear model for the data by the Akaike Information Criterion (AIC) using the **R** function `step()`. For this, perform a forward selection, backward elimination and stepwise selection in both directions. Compare the results.

   Then, compare the model you chose with the full model

   lm(R $\sim$ Age+S+Ed+Ex0+Ex1+M+N+U1+U2+W+X+poly(NW,3)+poly(LF,2))

   using the coefficient of determination $R^2$. Is this a reasonable comparison? If not, explain why and use another selection criterion to do the comparison.

b) Explain the statistical hypothesis test for the line "F-Statistic" in the summary of the linear model you chose in a), i.e. give $H_0$ and $H_1$ and the rejection rule. Then, explain what is given in the output and what conclusions can be drawn from that if we set $\alpha = 0.05$.

**Problem 3 (Additional)** The $i$-th observation is removed from the design matrix $X \in \mathbb{R}^{n\times p}$ of full rank and a response vector $\mathbf{Y}$ of a linear model. Without loss of generality, we assume $i = n$. Let $\mathbf{x}_n^\top$ denote the $n-$th row of $X$ and let

$$X = \begin{bmatrix} X_{-n} \\ \mathbf{x}_n^\top \end{bmatrix} \text{ and } \mathbf{Y} = \begin{bmatrix} \mathbf{Y}_{-n} \\ y_n \end{bmatrix},$$

where $X_{-n} \in \mathbb{R}^{(n-1)\times p}$ and $\mathbf{Y}_{-n} \in \mathbb{R}^{n-1}$. Denote the least squares estimator of the model with all observations by $\hat{\boldsymbol{\beta}}$ and write $\hat{\boldsymbol{\beta}}_{-n}$ for the least squares estimator when the $n-$th observation is removed.

Use Sherman-Morrison-Woodbury's Theorem to prove the following assertions:

a) Choose $A$, $\mathbf{u}$ and $\mathbf{v}$ of Equation (1) appropriately to show that

$$(X_{-n}^\top X_{-n})^{-1} = (X^\top X)^{-1} + \frac{1}{1-h_{nn}} \left[(X^\top X)^{-1}\mathbf{x}_n\mathbf{x}_n^\top (X^\top X)^{-1}\right],$$

where $h_{nn} < 1$ is th $n$-th diagonal entry of the hat matrix.

b) Use the assertion from part a) to show that

$$\hat{\boldsymbol{\beta}}_{-n} = \hat{\boldsymbol{\beta}} - \frac{r_n}{1-h_{nn}} \left(X^\top X\right)^{-1} \mathbf{x}_n .$$

c) Conclude that

$$y_n - \mathbf{x}_n^\top \hat{\boldsymbol{\beta}}_{-n} = \frac{1}{1-h_{nn}} r_n .$$

**Sherman-Morrison-Woodbury's Theorem** Let $A \in \mathbb{R}^{p\times p}$ be invertible and $\mathbf{u}, \mathbf{v} \in \mathbb{R}^p$. If $\mathbf{v}^\top A^{-1}\mathbf{u} \neq 1$ then

$$(A - \mathbf{u}\mathbf{v}^\top)^{-1} = A^{-1} + \frac{A^{-1}\mathbf{u}\mathbf{v}^\top A^{-1}}{1 - \mathbf{v}^\top A^{-1}\mathbf{u}}. \tag{1}$$

**Problem 4 (Additional)** We want to understand the orthogonalization in the function `poly` by doing the following. We consider the variable `LF` from the dataset `USCRIME`.

The polynomials that are used to calculate the transformed design matrix are defined by

$$\begin{aligned}
&F_0(x) = 1/\sqrt{n_2}\\
&F_1(x) = (x - a_1)/\sqrt{n_3}\\
&F_d(x) = \left[(x - a_d) * \sqrt{n_{d+1}} * F_{d-1}(x) - n_{d+1}/\sqrt{n_d} * F_{d-2}(x)\right]/\sqrt{n_{d+2}},
\end{aligned}$$

where $a_d$ and $n_d$ are given in `alpha` and `norm2` of `attributes(poly())$coefs` and $d$ is the degree of the polynomial.

a) Calculate the polynomials and compare them to the output of `poly()`.

b) Show that the resulting columns in the design matrix based on the output of `poly(LF,2)` are really orthogonal.

c) Show that the fitted values of `lm(R ~ poly(LF,2))` and `lm(R ~ poly(LF,2),raw=TRUE)` are equal. Also compare the model properties $R^2$ and AIC.