Statistical Science	HDDA bonus exercise 2
Leiden University	September 24, 2024

Follow the instructions below carefully:

- Do not load any package but the default ones (like `base`).
- Load the `assignmentA_grpB.Rdata` file with the assignment and group numbers substituted for `A` and `B`, respectively, into `R`. The file is available via Brightspace. Verify that you are using the correct file!
- Verify that two objects, named `X` and `Y`, have been loaded into `R`'s memory. These objects contain the data on the response vector (`Y`) and design matrix (`X`).
- Consider the linear regression model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ without (!) intercept and $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 \mathbf{I}_{nn})$, to explain the variation in the response `Y` by a linear combination of the columns of the design matrix `X`.
- Evaluate the generalized ridge estimator of regression parameter $\boldsymbol{\beta}$ of the linear regression model with target $\boldsymbol{\beta}_0$ equal to the zero vector with the last $p - \mathtt{B}$ elements replaced by one and penalty parameter $\boldsymbol{\Delta} = (1 - \rho)\mathbf{I}_{pp} + \rho \mathbf{1}_{pp}$ with $\rho = 1/(41 + \mathtt{B})$ in which the group number is substituted for `B`.
- Estimate the error variance σ^2 by $\frac{1}{n}\|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \boldsymbol{\Delta})\|_2^2$, where n denotes the sample size and $\hat{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \boldsymbol{\Delta})$ the generalized ridge estimator with $\boldsymbol{\beta}_0$ and $\boldsymbol{\Delta}$ as above. This yields part one of the solution of the bonus exercise that is to be send in.
- In the remainder replace $\boldsymbol{\Delta}$ by $\lambda\boldsymbol{\Delta}$ with $\lambda \in \{10, 20, 30, \ldots, 200\}$. Evaluate the mean squared error (MSE) of the obtained generalized ridge regression estimator, defined as:

$$\mathrm{MSE}[\hat{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \boldsymbol{\Delta})] \quad = \quad \mathrm{tr}\{\mathrm{Var}[\hat{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \boldsymbol{\Delta})]\} + \{\mathbb{E}[\hat{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \boldsymbol{\Delta})] - \boldsymbol{\beta}\}^\top \{\mathbb{E}[\hat{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \boldsymbol{\Delta})] - \boldsymbol{\beta}\},$$

  in which the true regression parameter $\boldsymbol{\beta}$ is assumed to equal the zero vector and the error variance $\sigma^2 = 1$. Find the λ that minimizes the MSE of $\hat{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \lambda\boldsymbol{\Delta})$. This minizing λ is part two of the solution that is to be send in.
- Send your answer in before `23:59` CET, October 7, 2024. Instructions for composing the email can be found in the `pdf`-file with information on the hand-in assigments.