MA8701 Advanced methods in statistical inference and learning

L10: Shrinkage methods for the GLM

Mette Langaas

2/6/23

Table of contents

1 Before we begin

1.1 Literature

- [ELS] The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics, 2009) by Trevor Hastie, Robert Tibshirani, and Jerome Friedman. Ebook. Chapter 4.4.1-4.4.3 (4.4.4 is covered in 3.2 of HTW).
- [HTW] Hastie, Tibshirani, Wainwrigh: "Statistical Learning with Sparsity: The Lasso and Generalizations". CRC press. Ebook. Chapter 3.2,3.7

and for the interested student

- [WNvW] Wessel N. van Wieringen: Lecture notes on ridge regression Chapter XXX

Supplemental sources useful for week 6:

- Bootstrap confidence intervals
- Short note on multiple hypothesis testing in TMA4267 Linear Statistical Models, Kari K. Halle, Øyvind Bakke and Mette Langaas, March 15, 2017.

---

pivec	0.10	0.20	0.30	0.40	0.5	0.6	0.70	0.8	0.9
odds	0.11	0.25	0.43	0.67	1.0	1.5	2.33	4.0	9.0

2 Generalized linear models

(HTW 3.1, 3.2, and TMA4315 GLM background)

2.1 The model

The GLM model has three ingredients:

1) Random component
2) Systematic component
3) Link function

We look into that for the normal and binomial distribution - to get multiple linear regression and logistic regression.

- Write in class
- Poll on standardization and centering.

2.2 Explaining β in logistic regression

- The ratio $\frac{P(Y_i=1)}{P(Y_i=0)} = \frac{\pi_i}{1-\pi_1}$ is called the *odds*.
- If $\pi_i = \frac{1}{2}$ then the odds is 1, and if $\pi_i = \frac{1}{4}$ then the odds is $\frac{1}{3}$. We may make a table for probability vs. odds in R:

- Odds may be seen to be a better scale than probability to represent chance, and is used in betting. In addition, odds are unbounded above.

We look at the link function (inverse of the response function). Let us assume that our linear predictor has k covariates present

$$
\begin{aligned}
\eta_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} \\
\pi_i &= \frac{\exp(\eta_i)}{1+\exp(\eta_i)} \\
\eta_i &= \ln(\frac{\pi_i}{1-\pi_i}) \\
\ln(\frac{\pi_i}{1-\pi_i}) &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} \\
\frac{\pi_i}{1-\pi_i} &= \frac{P(Y_i=1)}{P(Y_i=0)} = \exp(\beta_0) \cdot \exp(\beta_1 x_{i1}) \cdots \exp(\beta_k x_{ik})
\end{aligned}
$$

We have a *multiplicative model* for the odds.

So, what if we increase x_{1i} to $x_{1i}+1$?

If the covariate x_{1i} increases by one unit (while all other covariates are kept fixed) then the odds is multiplied by $\exp(\beta_1)$:

$$
\begin{aligned}
\frac{P(Y_i=1 \mid x_{i1}+1)}{P(Y_i=0) \mid x_{i1}+1)} &= \exp(\beta_0) \cdot \exp(\beta_1(x_{i1}+1)) \cdots \exp(\beta_k x_{ik}) \\
&= \exp(\beta_0) \cdot \exp(\beta_1 x_{i1}) \exp(\beta_1) \cdots \exp(\beta_k x_{ik}) \\
&= \frac{P(Y_i=1 \mid x_{i1})}{P(Y_i=0 \mid x_{i1})} \cdot \exp(\beta_1)
\end{aligned}
$$

This means that if x_{i1} increases by 1 then: if $\beta_1 < 0$ we get a decrease in the odds, if $\beta_1 = 0$ no change, and if $\beta_1 > 0$ we have an increase. In the logit model $\exp(\beta_1)$ is easier to interpret than β_1.

The response function as a function of the covariate x and not of η. Solid lines: $\beta_0 = 0$ and β_1 is 0.8 (blue), 1 (red) and 2 (orange), and dashed lines with $\beta_0 = 1$.