Bayesianische Regression

lineare und logistische Modelle

Lona Koers

LMU

25. Juli 2025

Outline

1. Motivation und Intuition
2. Bayesianische lineare Modelle
3. Bayesianische **generalisierte** lineare Modelle
4. Posterior Inference

Motivation und Intuition

TODO: gutes Beispiel

- Generalisierte Lineare Modelle (GLMs)
- Punktvorhersage vs. Verteilung vorhersagen
- warum reicht uns ein CI / PI

Outline

1. Motivation und Intuition
2. Bayesianische lineare Modelle
3. Bayesianische **generalisierte** lineare Modelle
4. Posterior Inference

Frequentistisches → bayesianisches lineares Modell

Annahmen:

1. i.i.d. Daten $\boldsymbol{D} = (\boldsymbol{y}, \boldsymbol{X})$
2. Kondition auf $\boldsymbol{X}$ (implizit)

Frequentistisches lineares Modell: $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\theta}, \sigma^2 \boldsymbol{I})$

3. Gewichtsparameter $\boldsymbol{\theta}$ als Zufallsvariable interpretieren

Bayesianisches lineares Modell:

$$\boldsymbol{y} \mid \boldsymbol{\theta}, \sigma^2 \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\theta}, \sigma^2 \boldsymbol{I})$$

Modelldefinition (Prior-Verteilungen)

Prior-Annahme für $\boldsymbol{\theta}$ (und evtl. σ^2) notwendig $\rightarrow$ sehr vielseitige Modell-Anpassung möglich

1. Normal-Invers-Gamma Prior:

$$\begin{aligned} \boldsymbol{\theta} \mid \sigma^2 &\sim \mathcal{N}(\breve{\boldsymbol{\mu}}, \sigma^2 \breve{\Sigma}) \\ \sigma^2 &\sim \mathsf{IG}(\breve{a}, \breve{b}) \\ \boldsymbol{\theta}, \sigma^2 &\sim \mathsf{NIG}(\breve{\boldsymbol{\mu}}, \sigma^2 \breve{\Sigma}, \breve{a}, \breve{b}) \end{aligned}$$

mit Prior Parametern: $\breve{\boldsymbol{\mu}}, \breve{\Sigma}, \breve{a}$ und $\breve{b}$

Vorteil: NIG-Prior ist mit Normalverteilungs-Likelihood konjugiert $\rightarrow$ exakte Inferenz möglich (mehr dazu später)

TODO: Bild

Uninformative Prior als Spezialfall der NIG-Prior

2. Uninformative Prior

z.B. mit NIG-Prior mit Prior Parametern

$$\breve{\boldsymbol{\mu}} = \mathbf{0}, \quad \breve{\Sigma}^{-1} = \mathbf{0} \text{ i.e., } \breve{\Sigma} \to \infty$$
$$\breve{a} = -\frac{p}{2}, \quad \breve{b} = 0$$

$\implies$ flache (und damit uninformative) Prior und maximaler Einfluss der Daten auf die Posterior:

$$\boldsymbol{\theta} \mid \sigma^2 \overset{a}{\sim} \mathcal{N}(\breve{\boldsymbol{\mu}}, \sigma^2 \infty) \implies p(\boldsymbol{\theta} \mid \sigma^2) \propto 1$$

TODO: Bild

Regularisierung: frequentistisch vs. bayesianisch

Erinnerung: *frequentistische* Regularisierung durch Minimierung von

$$\text{PLS}(\boldsymbol{\theta}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta})^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}) + \lambda\ \text{pen}(\boldsymbol{\theta})$$

mit Regularisierungs-Parameter $\lambda > 0.\backslash$

Bayesianische Regularisierung durch Wahl der Prior-Verteilung für $\boldsymbol{\theta}$

Regularisierung durch Prior Wahl

3. Ridge Regularisierung

Frequentistisch: $\text{pen}(\boldsymbol{\theta}) = \|\boldsymbol{\theta}\|_2^2$ TODO quote
Bayesianisch: $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \tau^2 \boldsymbol{I})$ mit $\tau^2 \propto \frac{1}{\lambda}$
TODO quote

4. Lasso Regularisierung

Frequentistisch: $\text{pen}(\boldsymbol{\theta}) = \|\boldsymbol{\theta}\|_1$
Bayesianisch:

$$\boldsymbol{\theta} \mid \boldsymbol{\tau}^2 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\tau}^2 \boldsymbol{I})$$

$$\tau_j^2 \overset{\text{i.i.d.}}{\sim} \text{Exp}(0.5\lambda^2), \quad j = 1, \dots, p$$

Problem: keine Variablenselektion (im Gegensatz zu frequentistischem Lasso)

$\rightarrow$ Alternative Priors für Variablenselektion: Spike and Slab, Horseshoe, u.v.m.

Regularisierung in Anwendung

Vorteile von bayesianischer Regularisierung sind u.a.:

- Probabilistisches Modell trotz Regularisierung
- Regularisierungparameter muss nicht als Hyperparameter optimiert werden (z.B. durch Prior auf τ^2)
- Mehr Anpassungsmöglichkeiten durch Prior-Spezifikation

TODO: Bild regularization Priors + update

Outline

1. Motivation und Intuition
2. Bayesianische lineare Modelle
3. Bayesianische **generalisierte** lineare Modelle
4. Posterior Inference

Bayesianisches LM $\rightarrow$ GLM (alt: Generalisierung des bayesianischen linearen Modells)

Outline

1 Motivation und Intuition

2 Bayesianische lineare Modelle

3 Bayesianische **generalisierte** lineare Modelle

4 Posterior Inference

Posterior Inference

Erinnerung: Bayes-Regel

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}) = \frac{p(\boldsymbol{y} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})}{\int p(\boldsymbol{y} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta}) d\boldsymbol{\theta}},$$

wobei $p(\boldsymbol{y} \mid \boldsymbol{\theta})$ die Modell-Likelihood ist.

bayesianisches LM: exakte Inferenz mit konjugierten Prioris

bayesianisches GLM: approximative Inferenz

Metropolis Hastings (MCMC)

Idee: Approximation der Posterior mit Markov chain Monte Carlo - was muss man für Regression anpassen?

Laplace Approximation (LA)

Idee: Approximation der Posterior mit einer Normalverteilung

- Mittelwert und Varianz werden mit IWLS berechnet

PPD

Literatur Empfehlungen

Falls euch das Thema interessiert und ihr mehr wissen wollt, kann ich folgende Literatur empfehlen:

- Bayesianische Regression (v.a. für praktische Anwendung): Gelman et al. [3]
- Prior Verteilungen (v.a. Shrinkage): Erp, Oberski, and Mulder [2] und Celeux et al. [1]
- Software: z.B. `brms` in `R`, `PyMC` in `python`

Referenzen I

[1] Gilles Celeux et al. "Regularization in Regression: Comparing Bayesian and Frequentist Methods in a Poorly Informative Situation". In: *Bayesian Analysis* 7.2 (June 2012), pp. 477–502. ISSN: 1936-0975, 1931-6690. DOI: 10.1214/12-BA716.

[2] Sara van Erp, Daniel L. Oberski, and Joris Mulder. "Shrinkage priors for Bayesian penalized regression". In: *Journal of Mathematical Psychology* 89 (Apr. 2019), pp. 31–50. ISSN: 0022-2496. DOI: 10.1016/j.jmp.2018.12.004.

[3] Andrew Gelman et al. *Bayesian Data Analysis*. 3rd ed. New York: Chapman and Hall/CRC, 2013. ISBN: 978-0-429-11307-9. DOI: 10.1201/b16018.