Computational Statistics and Data Analysis (MVComp2)

Exercise 7

Lecturer Tristan Bereau

Semester Wi23/24
Due Dec. 7, 2023, 23:59

1 Poisson regression (2 points)

Consider a response variable defined on the positive integer domain, $y_n \in \{0, 1, \dots\}$. We propose to fit a model using Poisson regression, such that the distribution's parameter $\lambda_n = \lambda_n(\boldsymbol{w}^\top \boldsymbol{x}_n)$ is a linear function of the input variables.

(a) Show that you can write Poisson regression as a generalized linear model (GLM).

(b) Use the GLM to determine the first two moments.

2 Binary-output linear regression (3 points)

Suppose we have binary input data, $x_i \in \{0, 1\}$ and output two-dimensional response vector, $y_i \in \mathbb{R}^2$. The data is the following

x	y
0	$(-1,-1)^\top$
0	$(-1,-2)^\top$
0	$(-2,-1)^\top$
1	$(1,1)^\top$
1	$(1,2)^\top$
1	$(2,1)^\top$

Embed each x_i into two dimensions using the following basis function

$$\phi(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \phi(1) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The model becomes $\hat{y} = \boldsymbol{W}\phi(x)$, where $\boldsymbol{W}$ is a 2×2 matrix. Compute the maximum likelihood estimator for $\boldsymbol{W}$.

3 Posterior credible interval (3 points)

The Bayesian analog of a confidence interval is called a credible interval. Let's work with that here. Consider $X \sim \mathcal{N}(\mu, \sigma^2 = 4)$. The mean, μ, is unknown, but has a prior $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2 = 9)$. After seeing n samples the posterior is $\mu \sim \mathcal{N}(\mu_n, \sigma_n^2)$.

(a) Determine μ_n and σ_n^2.

(b) How big does n have to be to ensure

$$p(a \leq \mu_n \leq b | D) \geq 0.95,$$

where (a, b) is an interval centered on μ_n of width 1 and D is the data?

Hint: 95% of the probability mass of a Gaussian is within $\pm 1.96\sigma$ of the mean.

4 Integration by Monte Carlo (2 points)

Estimate

$$\ell = \int_0^1 \mathrm{d}x \int_0^1 \mathrm{d}y \, \frac{\sin(x) \mathrm{e}^{-(x+y)}}{\ln(1+x)}$$

via Monte Carlo, and give a 95% confidence interval.