Bayesian logistic regression

Blaž Erzar

Machine Learning For Data Science 1 2023/24
Homework 6 report

1 Introduction

In this homework, we use the Bayesian inference to infer the relationship between the input variables. We are working on the basketball shot dataset with two features, **angle** and **distance**. For the inference, we use MCMC, and Laplace approximation at the end.

2 Results

Before looking at the results, we conjecture that the coefficient corresponding to the distance feature will be distributed normally with the mean $\mathbb{E}[\beta_{\text{distance}}] = -1.6$, see Figure 1. For all computations, we normalize the features between 0 and 1 because the angles and distances are positive. We use standard normal priors and generate 4 chains of length 10 000 with 1 000 samples as burn-in.

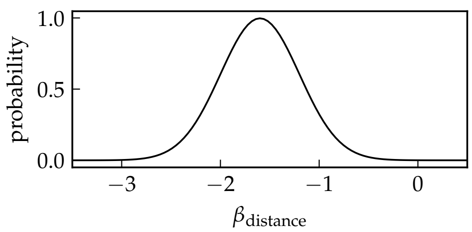

Figure 1: *Personal opinion about the coefficient $\beta_{distance}$.*

After looking at the results in Table 1, we can see that the coefficient is indeed negative, but not as much and with a smaller variance. For both cases, we also show the contour plot in Figure 2. For the Laplace approximation, the results are similar to MCMC.

Approach	β_0	β_{angle}	β_{distance}
MCMC	0.565 ± 0.145	-0.114 ± 0.179	-1.439 ± 0.192
Sample	0.306 ± 0.505	-0.492 ± 0.645	-0.135 ± 0.726
Laplace	0.562 ± 0.146	-0.111 ± 0.181	-1.436 ± 0.191

Table 1: *Statistics for 10 000 generated samples using MCMC with the whole dataset, using MCMC with a subset of 50 rows and Laplace approximation on the whole dataset. We show the sample mean and standard deviation.*

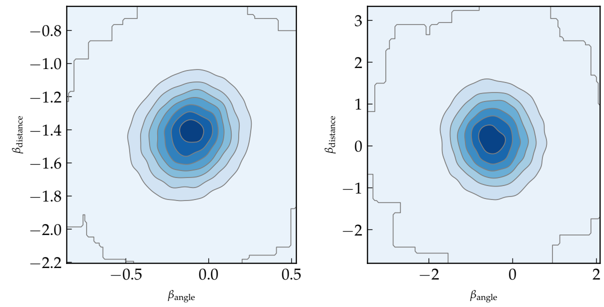

Figure 2: *Contour plots of samples generated using MCMC using the whole dataset (left) and a subset of 50 rows (right).*

To fit the Gaussian to the posterior for the Laplace approximation, we need to approximate the mean and the variance. For the mean, we take the maximum of the posterior (in log space for easier computation), for which we need its gradient:

$$\frac{\partial}{\partial \beta} \log p(\beta|y) = \sum (y_i - \sigma(\mathbf{x}_i^\top \beta))\mathbf{x}_i^\top - (\beta - \mu)^\top \Sigma^{-1}.$$

Variance can then be computed from the Fisher information using curvature:

$$\frac{\partial^2}{\partial \beta^\top \partial \beta} \log p(\beta|y) = -\sum \sigma(\mathbf{x}_i^\top \beta)(1 - \sigma(\mathbf{x}_i^\top \beta))\mathbf{x}_i \mathbf{x}_i^\top - \Sigma^{-1}.$$

3 Discussion

When using the whole dataset, the distance is more important for missing a shot, which is expected for most players. This question can also be answered probabilistically by computing

$$P(\beta_{\text{distance}} < \beta_{\text{angle}}) = 1.0.$$

When using only a small sample, the uncertainty is much higher and because we are working with only a single concrete sample, the results are vastly different. For this sample, the angle is more important for not making a shot, while the distance is only slightly correlated. Generating another sample would probably change the results further.

Another probabilistic question we answer is whether a higher angle decreases shot accuracy. Computing the probability

$$P(\beta_{\text{angle}} < 0) = 0.737,$$

we can see that in most cases it does.