

Bayesian Logistic Regression

Aljaž Konec

I. INTRODUCTION

In this report we showcase the implementation of Bayesian logistic regression. First we compare the posterior distributions of the coefficients for all shots and subsampled shots. Then we discuss the importance of the features and how they are linked to the probability of making a shot.

II. PERSONAL OPINION

In our personal opinion, distance from the basket plays a more important role in the probability of making a shot than the angle of the shot. The further away from the basket the player is, the less likely the shot is made. Therefore we expect the coefficient β_1 to have mostly negative values. Since players train on shooting there will be certain distances where the player is most comfortable shooting and therefore the coefficient β_1 will have a positive value. All of this is summarized in Figure 1.

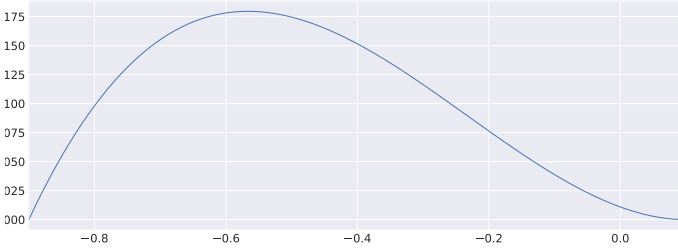


Figure 1: Personal opinion on the coefficient for Distance.

III. RESULTING POSTERIOR DISTRIBUTIONS

We perform MCMC sampling on all shots and 50 randomly subsampled shots to obtain two posterior distributions for both coefficients. The results are shown in Figure 2. Comparing the posterior distributions to our personal opinion we can see that the distance coefficient is in fact negative just not to the extent we expected. Inferring from subsampled data results in similar distributions where the distance coefficient is negative and the angle coefficient is close to zero.

Comparing the distributions for all shots and subsampled shots it can be observed that the former gives a much smaller area of probable coefficient values. This indicates that the model is more certain about the coefficients when using more data. Table I shows the mean and standard deviation of the posterior distributions for both coefficients. The standard deviation for subsampled distribution is much larger, which again shows the larger uncertainty in the model when using less data.

	All Data	Subsampled Data
Angle Coefficient	-0.0015 ± 0.0021	-0.0068 ± 0.0126
Distance Coefficient	-0.1527 ± 0.0201	-0.3758 ± 0.1307

Table I: Means and standard deviations of both posterior distributions for the angle and distance coefficients.

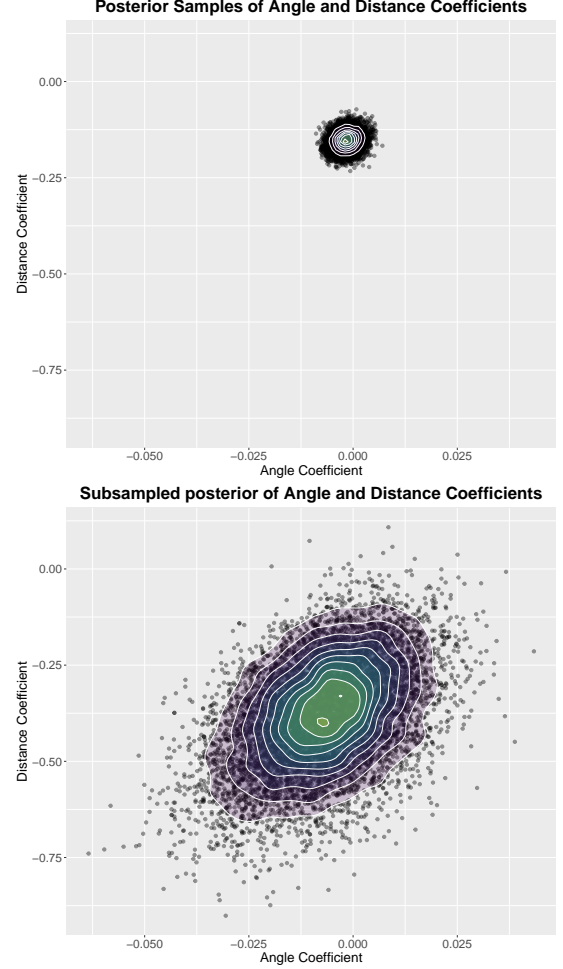


Figure 2: Posterior distributions for angle and distance coefficients for all shots and subsampled shots.

IV. IMPORTANCE OF FEATURES

From the posterior distribution in Figure 2 we can see that the angle coefficient values are distributed very closely around zero while the distance coefficient is distributed at -0.1527 ± 0.0201 . Since the distance coefficient is (in absolute terms) much larger than the angle coefficient, we can conclude that the distance from the basket is far more important than the angle of the shot.

V. SHOT SUCCESS BASED ON INCREASING ANGLE

We can calculate this by calculating the probability that the angle coefficient is negative, indicating that shot success decreases with increasing angle. The probability can be calculated as the proportion of posterior samples that are negative:

$$P(\beta_0 < 0) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{(x < 0)}(\beta_0^{(i)}) = 0.7627 \quad (1)$$

We conclude that the coefficient of the angle is negative 76.27% of the time, indicating that the shot success decreases with increasing angle.