Bayesian Logistic Regression

Ivan Nikolov[1]

[1] *in7357@student.uni-lj.si, 63190378*

Coefficient Priors

What is your personal opinion about the coefficient beta for distance?
Based on my limited basketball knowledge, I assumed that distance will negatively affect the probability of making a shot. That means the bigger the distance, the less likely it is that the player will score. Based on that, I set a prior $\mathcal{N}(-1, 2.5)$. The default prior standard deviation was 2.5, so I only changed the mean.

For the angle, I also assumed that large angles will negatively impact the shot success, but not in the same magnitude as the distance. Based on this, I set the prior as a $\mathcal{N}(-0.5, 2.5)$.

I left the prior on the intercept as the default $\mathcal{N}(0, 2.5)$.

Results

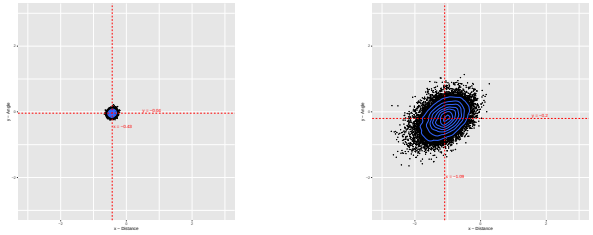
As a part of preprocessing, I standardized both features so that they have a mean of 0 and a standard deviation of 1.

For the posterior I used a MCMC approximation with 20000 samples. The ESS of the model build on the whole dataset for the intercept, angle, and distance were: 19912, 19047, 19751 accordingly. When using only 50 samples, they somewhat dropped to 16133 (intercept), 15378 (angle) and 14845 (distance).

On Figure 1 a scatter plot from the posterior samples of the coefficients using the whole dataset and 50 samples are shown. It is clearly visible that the variance is much smaller when using the whole dataset. The priors also had a greater effect on the distribution when only 50 samples were used (the distributions are shifted to the right). This is also visible in the confidence intervals in Table 1, and the density estimates on Figure 2.

Parameter	Model	Median	95% CI
Angle	50	-0.19	[-0.89, 0.44]
Angle	Whole	-0.04	[-0.15, 0.07]
Distance	50	-1.08	[-1.86, -0.39]
Distance	Whole	-0.43	[-0.54, -0.32]

Table 1. Parameter Estimates with 95% Confidence Intervals

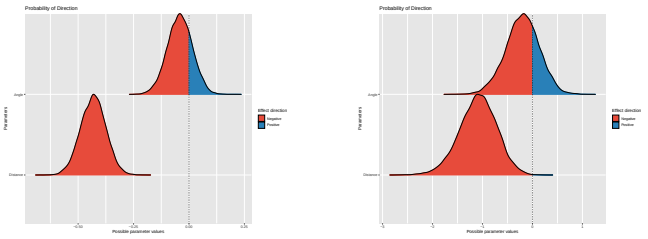


(a) Whole dataset **(b)** 50 random samples

Figure 1. Scatter plots on posterior samples. The dashed red lines represent the means.

Which is more important for shot success, angle or distance? We can calculate this as the probability of the distance coefficient magnitude being bigger than the angle coefficient (in absolute values). That is, $P(|\beta_2| - |\beta_1| > 0) = 1$ (β_2 - distance β_1 - angle). Based on this, the distance will always have more impact than the angle.

Does shot success increase or decrease with increasing angle (the further on the sides we are)? I answered this by calculating the probability of $P(\beta_1 < 0) = 0.76$. This means that the angle coefficient is more likely to be negative and decrease the probability of the shot success in our model.



(a) Whole dataset **(b)** 50 random samples

Figure 2. Posterior of logistic regression coefficients.