

Logistic Regression

Aljaž Konec, 63190019

I. INTRODUCTION

Logistic regression is one of the most widely used machine learning algorithms for binary classification. Its use can also be expanded to work on multi-class and ordinal classification problems. Multi-class regression, also known as multinomial logistic regression, extends binary classification to allow for K classes to be predicted. Ordinal regression, on the other hand, is used when the classes have a natural, clear order to them. Some such examples may include grades in school, customer satisfaction and modeling of human levels of preference (a scale from, say, 1-5 for "very poor" through "excellent"). In this report we first showcase the implementation details of multinomial logistic regression as well as ordinal logistic regression. The next section describes how the coefficients of multinomial regression are related to the target classes. In the final section we showcase an example of a data generating process that is better modeled for ordinal regression than for multinomial regression.

II. MODEL IMPLEMENTATION

Multinomial logistic regression

The multinomial logistic regression model differs from the binary logistic regression model in that it has K sets of coefficients, one for each class. To calculate the probability of a sample belonging to a certain class, we use the softmax function:

$$P(y = k|x) = \frac{e^{\beta_k^T x}}{\sum_{j=1}^K e^{\beta_j^T x}} \quad (1)$$

where β_k is the set of coefficients for class k . The loss function is then defined as the negative log-likelihood over all samples:

$$-\log L(\beta) = -\sum_{i=1}^N \sum_{k=1}^K \delta_{ik} \log P(y = k|x_i) \quad (2)$$

where δ_{ik} is the indicator function that is 1 if $y_i = k$ and 0 otherwise. To account for the intercept term, we add a column of ones to the input matrix X . We also add L2 regularization to the loss function. The resulting β is then a matrix of shape (K, M) where M is the number of features in the input matrix X . For predicting the class of a sample, we simply take the class with the highest probability for that sample.

Ordinal logistic regression

Ordinal logistic regression is a generalization of binary logistic regression to multiple classes that have a natural order to them. We train $K - 2$ thresholds τ that separate the classes. These thresholds have to be ordered, i.e. $\tau_1 < \tau_2 < \dots < \tau_{K-2}$. We achieve this by appending Δ_{k-1} to the vector of coefficients β for each threshold τ_k . And defining $\tau_0 = -\infty \leq \tau_1 = 0 \leq \tau_2 = \tau_1 + \Delta_1 \leq \dots \leq \tau_{K-1} = \tau_{K-2} + \Delta_{K-2} \leq \tau_K = \infty$. The probability vector of size K is then defined (component-wise) as:

$$\begin{aligned} P(y = k|x) &= P(y \leq k-1|x) - P(y \leq k|x) \\ &= \frac{1}{1 + e^{\tau_k - \beta_k^T x}} - \frac{1}{1 + e^{\tau_{k-1} - \beta_{k-1}^T x}} \end{aligned}$$

Calculating the loss function is done in a similar way as for multinomial regression, but with the probability vectors defined above. The resulting β is then a matrix of shape $(K - 1, M)$ where M is the number of features in the input matrix X .

III. INTERPRETATION OF COEFFICIENTS

To better understand the coefficients of the multinomial logistic regression model we train a classification model on a dataset of Basketball Shot Types [1]. The dataset contains 8 features: Shot Type, Competition, Player Type, Transition, Two-Legged, Movement, Angle and Distance. The target variable is the Shot Type, which has 6 classes: Above Head, Dunk, Hook Shot, Layup, Tip-in and Other. Competition, Player Type and Movement are categorical features that were one-hot encoded. All features were then centered and scaled appropriately. This resulted in 12 features and one target variable.

Since coefficients are an estimate they contain some uncertainty. To model this uncertainty we use 100 repetitions of bootstrapping. The results are shown in Figure 1. A positive coefficient means that the feature is positively correlated with the target class. It can be seen that Transition, Angle and Player Type have coefficients that are close to zero. This makes intuitive sense as scoring a basket is not dependent on whether the shot was taken in transition. The angle at which the shot was taken is also not important as only very rare shots are taken at extreme angles. All players practice shooting and therefore the player type does not have a significant impact on the shot type.

With increasing distance from the basket, Tip-ins and Dunks become less likely or even impossible. This is reflected in the very low negative coefficients for Distance in those classes. The classic Above Head shot is more likely to be taken from a distance and as such has a high positive coefficient for Distance. A similarly high Distance coefficient can be seen for Other shots.

Another very important feature is movement and no movement. An Above Head shot is very difficult to make while moving and as such has a very high negative coefficient for Movement. The opposite is true for Layups, which are often taken while moving. This is reflected in a positive coefficient for Drive Movement and a negative coefficient for No Movement. The Hook Shot is also more likely to hit while stationary, which is reflected in a high coefficient for No Movement.

Dunks and Tip-ins are easier to make if the player is taller and can jump higher. Therefore, these classes are negatively correlated with the Under 14 Competitions as they are played by young players who are not fully grown yet. Looking further into the Dunk class we can see that only the NBA competition has a positive coefficient associated with it while the other competitions have negative coefficients. From this we can gather that Dunking is more common in the NBA, which could be due to the fact that the NBA boasts the best players in the world and they are more likely to be able to successfully perform a dunk.

The TwoLegged feature is also important for many classes. Performing a Tip-in, Dunk or Layup is easier if the player can jump off of two legs and therefore the TwoLegged feature has a positive coefficient for those classes. The Other class has a mix

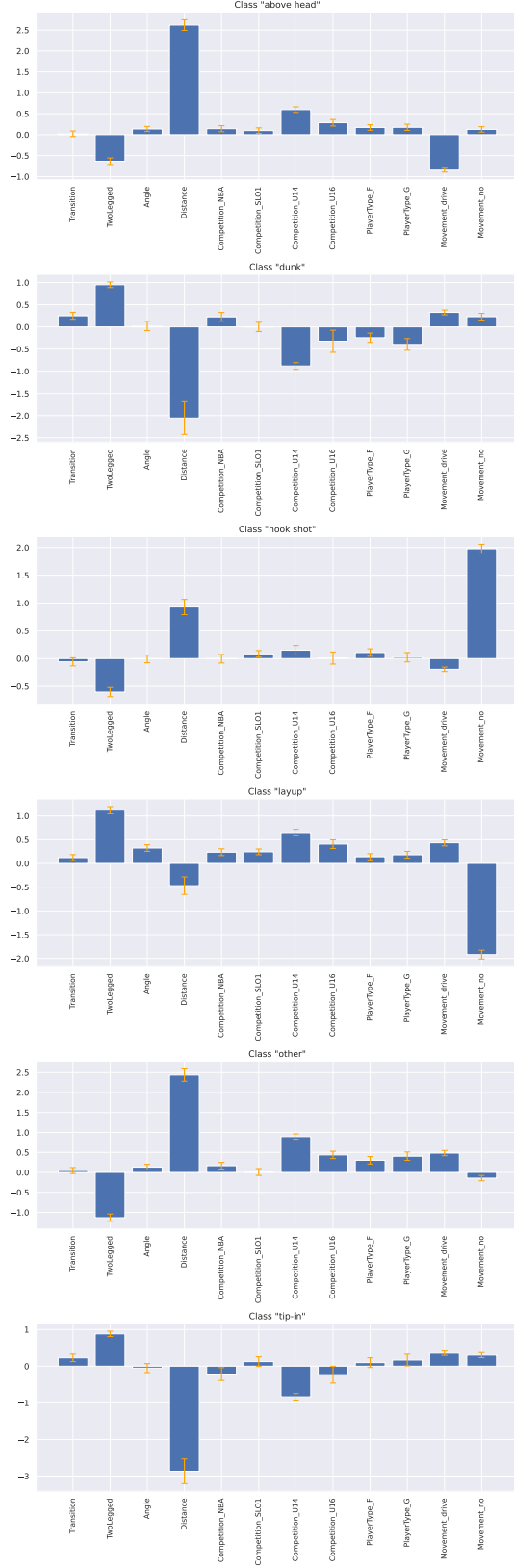


Figure 1: Coefficients of the multinomial logistic regression model.

of shots where the player might jump off of one or two legs and as such has a low coefficient for TwoLegged.

IV. OUTPERFORMANCE OF ORDINAL REGRESSION

To showcase the outperformance of ordinal regression over multinomial regression we created a new data generating process such that: The target variables are integers between 0 and 3, the feature matrix X contains 12 features, each drawn from a standard normal distribution. Generating a training set of 50 samples and a test set of 1000 samples we trained both a multinomial and an ordinal logistic regression model. The results are shown in Table I.

Model	Accuracy	Log-loss
Multinomial	0.323	1.708
Ordinal	0.339	1.433

Table I: Results of the data generating process.

We can see that Ordinal regression has a lower Log-loss value, this is due to the fact that extreme values in the features have higher chances to be classified either into 0 or 3 as these are the most extreme classes. While the multinomial regression model there is no such clear separation and such values can be classified into any of the 4 classes.

V. CONCLUSION

Data in the real world is rarely binary and as such multinomial regression models are needed to classify data. Likewise, ordinal regression models can be very useful for situations where modeling human preferences is needed. This report showcased that the implementation of both models is relatively simple and that the coefficients of the models can be interpreted to gain insight into the data.

REFERENCES

- [1] F. Erčulj and E. Štrumbelj, "Basketball shot types and shot success in different levels of competitive basketball," *PLOS ONE*, vol. 10, no. 6, pp. 1–14, 06 2015. [Online]. Available: <https://doi.org/10.1371/journal.pone.0128885>