Multinomial and Ordinal Logistic Regression Implementations

Teodora Taleska

1. Multinomial Logistic Regression

The primary task in this model is to model the probabilities of each class given the input features.

1.1 Model Structure

In the implementation, the model is fitted using maximum likelihood estimation. We begin by calculating the log-likelihood of the model parameters, which is given by:

$$\mathscr{L}(\beta) = \sum_{i=1}^{n} \log(P(y_i|X_i, \beta))$$

where $P(y_i|X_i, \beta)$ is the probability of class y_i for the i-th sample, computed using the *softmax* function.

The probabilities are computed via the softmax function:

$$P(y_i = j|X_i) = \frac{\exp(\beta_j^T X_i)}{\sum_{k=1}^{m} \exp(\beta_k^T X_i)}$$

where X_i is the feature vector, β_j is the coefficient for class j, and m is the number of classes. This ensures that the sum of the predicted probabilities for all classes is 1 for each observation.

1.2 Optimization and Coefficient Estimation

The model parameters, including the weights (β) and intercepts, are estimated using the L-BFGS-B optimization algorithm. The optimization maximizes the log-likelihood function by updating the parameters iteratively. The gradient of the log-likelihood with respect to the parameters is computed analytically to efficiently guide the optimization process.

$$\text{Gradient} = \frac{\partial \mathscr{L}(\beta)}{\partial \beta}$$

The gradient is calculated using the chain rule, with the softmax function applied to the logits (the linear combination of features and weights). This allows us to adjust the parameters in the direction of maximum likelihood.

2. Ordinal Logistic Regression

This model is used for classification tasks where the target variable has ordered categories (ordinal data). Unlike multinomial logistic regression, where the classes are independent, ordinal logistic regression assumes an inherent order between the categories.

2.1 Model Structure

In this model, we compute the probabilities of each class using a cumulative distribution function (CDF) based on the thresholds between classes. The model computes the probability of a sample belonging to a certain class using the logistic CDF:

$$P(y_i \leq j|X_i) = \frac{1}{1 + \exp(-(X_i^T \beta + \delta_j))}$$

where δ_j are the thresholds between classes. The final probability for class j is then the difference between the cumulative probabilities for class j and class $j-1$.

2.2 Threshold Calculation and Deltas

The thresholds between the classes are determined by cumulative deltas, which are optimized during the fitting process. The deltas control the shift between thresholds. We start by setting initial deltas and compute the thresholds based on the cumulative sum of deltas. These thresholds are then used to compute the class probabilities for each sample.

$$t_j = \sum_{k=1}^{j} \delta_k$$

The model optimizes these deltas to maximize the log-likelihood, with the objective being the same as in multinomial logistic regression: maximizing the probability of the correct class labels given the input features.

2.3 Optimization and Coefficient Estimation

The fitting procedure for ordinal logistic regression involves optimizing both the coefficients (β) and deltas using the L-BFGS-B algorithm. The negative log-likelihood function is minimized:

$$\mathscr{L}(\beta, \delta) = -\sum_{i=1}^{n} \log(P(y_i|X_i, \beta, \delta))$$

where the probabilities are computed based on the thresholds and coefficients, and the deltas are constrained to be positive.