Ordinal Regression

Rahul Ramachandran
cs21btech11049

Rishit D
cs21btech11053

October 6, 2023

Contents

1   Ordinal Regression: Summary

Ordinal random variables are based on classfifcation random variables but with a stochastic ordering. For instance, support for various political parties is considered a classfication problem whereas support for various parts of the political spectrum is an ordinal regression problem. Consider k categories which are the qualitative representation of the ordering. We would want to determine the probabilities of landing in various categories based on a set of independent features.

Assume the independent variables to be represented through $\mathbf{x}$ with the probability of being in category i as $\pi_i(\mathbf{x})$. Now define cumulative probabilities $\gamma_i(\mathbf{x})$ as follows:

$$\gamma_i(\mathbf{x}) = \sum_{a=1}^{i} \pi_a(\mathbf{x}) \tag{1}$$

Our ordinal regression model is based on the General Linearized Model, represented as follows:

$$g(\gamma_i(\mathbf{x})) = \theta_i - \beta^T \mathbf{x} \tag{2}$$

where θ_i represents the cut-point for category i, β represents a vector of unknown parameters, $\mathbf{x}$ represents the vector of independent features and $g(x)$ is the link function for the GLM. Note that link functions are monotone and map $(0, 1)$ to $(-\infty, \infty)$.

Intuitively, we note that our set of cut-point $\{\theta_i\}$ must be monotone. One can note that we obtain a set of parallel hyperplanes that separate out different categories when we consider our link function as the identity function. But based on the nature of the dataset given to us (in specific, the possible distribution from which the outcome Y may be generated), we choose different link functions and obtain parameters. One may also note that we only have a single set of unknown paramaters β common to all categories, thus geneereating the set of parallel hyperplanes.

Now consider two input vectors x_1 and x_2. Then we have,

$$g(\gamma_i(x_1)) - g(\gamma_i(x_2)) = \beta^T(x_2 - x_1) \tag{3}$$

We observe that the right hand side is completely independent of the category. Hence, by monotonicity of the link function g, we have that either of the following two equations to hold.

$$\gamma_i(x_1) > \gamma_i(x_2) \tag{4}$$
$$\gamma_i(x_1) < \gamma_i(x_2) \tag{5}$$

for all categories i.

1.1   Proportional Odds Model

Assume that odds of $\mathbf{x}$ belonging to a category i such that $i < j$ is modelled as:

$$T_j(\mathbf{x}) = C_j e^{-\beta^T \mathbf{x}} \tag{6}$$

This model is termed the proportional odds model as the ratio of odds of $i < j$ when input vectors are x_1 and x_2 is independent of the category.

$$\frac{T_j(x_1)}{T_j(x_2)} = e^{\beta^T(x_2 - x_1)} \tag{7}$$

2   Key Differences With Multiclass Classification

3   Likelihood & Parameter Estimation