CS3390 - Foundation of Machine Learning

Group: 98.

Assignment-1

Roll NO: CS21BTECH11023 Jarupula sai Kumar

Roll NO: CS21BTECH11008 Bhende Adarsh Suresh.

Q. (a) providing a brief summary of the Paper. Ordinal data is the kind of data which categorical with clear order or ranking for the categories. However, It is impossible to measure precisely how the categories differ from one another.

Therefore, we use Ordinal regression Approach to predict the Output values for this kinds of data. This method has similarities with both regression and classification. On some Continuous scale, these categories can be conceived of as Continuous intervals.

Now, Here we are discussing two widely used models for Ordinal regression.

(i) proportional Odds Model :- This model proportional to odds models specifies that the Odds for the Event $Y \leq j$ is given by.

$$K_j(x) = K_j \exp(-\beta^T x)$$

Here, if the data is binary Categorical data $k=1$, Then model is same as that of Linear Logistic model.

(ii) Proportional Hazards model :-

The probabilistic of surviving beyond time t for a given covariate x is given by

$$-\log(S(t;x)) = \Lambda_0(t)\exp(-\beta^T x).$$

Where, $\Lambda_0(t) = \int_0^t \lambda_0(s)\,ds.$

Let us Assume that there are K categories.

Let $\pi_1(x), \pi_2(x), ---\pi_K(x)$ be the probabilities of K classes respectively for the input value x.

We define $\gamma_j(x)$ to the probability of the Event $y \le j$.

$$\gamma_j(x) = \pi_1(x) + \pi_2(x) + \cdots + \pi_j(x)$$

$$\forall_{1 \le j \le K}.$$

$$\frac{\gamma_j(x)}{1-\gamma_j(x)} = k_j \exp(-\beta^T x).$$

$$\log\left(\frac{\gamma_j(x)}{1-\gamma_j(x)}\right) = \theta_j - \beta^T x.$$ where $\theta_j = \log(k_j)$

Now, We know that, Therefore Odds ratio is given by.

$$\frac{\gamma_j(x)}{1-\gamma_j(x)} = \exp(\theta_j - \beta^T x).$$

Multi-class Classification, the Odds for the Even

$$\frac{P(Y \le j)}{1-P(Y \le j)} = \frac{\sum_{i=1}^{j} \exp(w_i^T x)}{\sum_{i=1}^{K} \exp(w_i^T x)}.$$

'n' be the number of data points.

n_j be the number of points which belongs to category j.

$$\sum_{i=1}^{K} n_i = n$$

The likelihood is Given by

$$L = \pi_1(x)^{n_1} \pi_2(x)^{n_2} \cdots \pi_k(x)^{n_k}$$

$$L = (\gamma_1(x))^{n_1} (\gamma_2(x) - \gamma_1(x))^{n_2} \cdots (\gamma_k(x) - \gamma_{k-1}(x))^{n_k}$$

In case of Multi-class classification the likelihood is Given by.

$$\boxed{L = \frac{\exp\left(\sum_{i=1}^{k} n_i w_i^T x\right)}{\left(\sum_{i=1}^{k} \exp(w_i^T x)\right)^n}}$$

The Differences with Other regression problems:
The regression models the Cumulative probability of the Categories of Classes, Which is One of the key difference. It makes the assumption that the likelihood of falling into lower category or one that is Equal to a certain Category versus falling into a higher category is proportional to the values of Independent Variables unlike Other regression problems that deals with predicting the Output for the given datapoints using an objective function. Thus unlike Other regression issues, its nature is sigificantly affected by changing the sequence of the dataset.

② (b). The Parameter estimation technique derivation for Ordinal Likelihood

Let us Assume that there are K categories.

Let $\pi_1(x), \pi_2(x), \ldots \pi_K(x)$ be the probabilities of K classes respectively for the input value x.

We define $\gamma_j(x)$ to the probability of the Event $y \le j$.

$$\gamma_j(x) = \pi_1(x) + \pi_2(x) + \ldots + \pi_j(x)$$

$$\forall 1 \le j \le K.$$

$$\frac{\gamma_j(x)}{1-\gamma_j(x)} = k_j \exp(-\beta^T x).$$

$$\log\left(\frac{\gamma_j(x)}{1-\gamma_j(x)}\right) = \theta_j - \beta^T x. \qquad \text{Where } \theta_j = \log(k_j).$$

Now, We know that, Therefore Odds ratio is given by.

$$\frac{\gamma_j(x)}{1-\gamma_j(x)} = \exp(\theta_j - \beta^T x).$$

Multi-class Classification, the Odds for the Event

$$\frac{P(Y \le j)}{1-P(Y \le j)} = \frac{\sum_{i=1}^{j} \exp(w_i^T x)}{\sum_{i=1}^{K} \exp(w_i^T x)}$$

'n' be the number of data points

n_j be the number of points which belongs to category j.

$$\sum_{i=1}^{k} n_i = n$$

The likelihood is Given by

$$L = \pi_1(x)^{n_1} \pi_2(x)^{n_2} \cdots \pi_k(x)^{n_k}$$

$$L = (\gamma_1(x))^{n_1} (\gamma_2(x) - \gamma_1(x))^{n_2} \cdots (\gamma_k(x) - \gamma_{k-1}(x))^{n_k}$$

In Case of Multi-class Classification the likelihood is Given by.

$$\boxed{L = \frac{\exp\left(\sum_{i=1}^{k} n_i w_i^T x\right)}{\left(\sum_{i=1}^{k} \exp(w_i^T x)\right)^n}}$$

Explanation: The Ordinal likelihood function L represent a cummulative odd model in Ordinal regression, aiming to estimate threshold parameters $\{\gamma_1, \gamma_2, \gamma_3 \ldots \gamma_{k-1}\}$ that separates Ordered Categories. TO derive these parameters through likelihood Estimation, one computes the log-likelihood function, taking the natural logarithm of L. Which specifies the product term into a sum. Then, Numerical optimization methods, like gradient ascent, are employed to maximum the log. likelihood by iteratively adjusting the threasholds until convergence.