FUNDAMENTALS OF MACHINE LEARNING
ASSIGNMENT-1

DEEPSHIKHA-CS21BTECH11016
TEJAL-CS21BTECH11058

Question 2

Part(a):

(1) Summary : The paper delves into a class of regression models tailored for handling ordinal data. Of particular significance are two models: the Proportional Odds Model and the Proportional Hazards Model, known for their practical utility due to their straightforward interpretability.These linear models can be viewed as an extension of generalized linear models but adapted for multivariate analysis, making them versatile tools for a wide range of applications. Moreover, we explore extensions to nonlinear models within this framework. Remarkably, even in the realm of nonlinear models, the method of iteratively reweighted least squares demonstrates convergence to the maximum likelihood estimate. This property streamlines the computational aspects of these models, rendering them more accessible and applicable.The paper illustrates the practical utility of these models with real-world examples, showcasing how they can effectively capture and analyze the nuanced relationships present in ordinal data, thereby enhancing our ability to make meaningful inferences and predictions.

(2) (a) Likelihood in ordinal regression are distinct from those in multi-class classification due to the following reasons:
   (i) In ordinal regression, it is based on the cumulative probabilities.It estimates the probability that an observation falls into or above a specific category based on the predictor variables.
   (ii) The likelihood in multi-class classification is typically based on the probability of an observation belonging to each class. It estimates the probability of an observation belonging to each class independently, without considering any ordinal relationship among the classes.

   (b) Odds ratio in ordinal regression are distinct from those in multi-class classification due to the following reasons:
   (i) In ordinal regression, odds ratios are used to describe the effect of predictor variables on the odds of being in a higher category (or a specific category) compared to a reference category. These odds ratios provide information about the change in the odds as the predictor variables change.
   (ii) Odds ratios are less commonly used in multi-class classification because they are primarily employed when dealing with binary outcomes. Instead, multi-class models often provide class probabilities, and decision boundaries are determined based on these probabilities.

(3) (a) The difference between likelihood of regression and ordinal regression is:
   (i) The likelihood function in standard regression is typically based on the normal distribution (in linear regression) or other distributions appropriate for

Date: 8th October 2023.

continuous data. It estimates the probability of observing the continuous outcome values given the model parameters. The likelihood is a continuous probability density function.

(ii) The likelihood function in ordinal regression is tailored to handle ordinal data. It estimates the probability of an observation falling into or above a specific ordinal category given the predictor variables. The likelihood function in ordinal regression is based on cumulative probabilities and takes into account the ordinal nature of the outcome.

(b) The difference between odds ratio of regression and ordinal regression is:

(i) Odds ratios are not commonly used in standard regression, particularly in linear regression. Linear regression models provide coefficients that represent the change in the expected value of the continuous outcome for a one-unit change in a predictor variable. Odds ratios are more relevant in logistic regression, which is used for binary outcomes.

(ii) Odds ratios are a fundamental part of ordinal regression. They describe how changes in predictor variables impact the odds of an observation being in a higher ordinal category compared to a reference category. These odds ratios provide insights into the direction and magnitude of the effect of predictors on the ordinal outcome.

Part(b):

(1) The ordinal likelihood model is a multivariate generalized linear model, combining the linear systematic structure with multinomial variation.

(2) The parameter estimates can be obtained using an Iteratively Reweighted Least Squares approach which involves iteratively updating the parameter estimates to maximize the likelihood function.

(3) The contribution from a single multinomial observation $(n_1, ..., n_k)$ to the likelihood function is $\pi^{n_1} ... \pi^{n_k}$, with the probabilities π^j satisfying the link function.We define

$$\begin{aligned} R_1 &= n_1 & Z_1 &= R_1/n, \\ R_2 &= n_1 + n_2 & Z_2 &= R_2/n \\ &\vdots & &\vdots \\ R_k &= \Sigma n_j & Z_k &= R_k/n = 1 \end{aligned}$$

(4) In terms of the parameters of the cumulative transformation, the likelihood can be written as the product of k-1 quantities,

$$\left\{ \left(\frac{\gamma_1}{\gamma_2}\right)^{R_1} \left(\frac{\gamma_2 - \gamma_1}{\gamma_2}\right)^{R_2 - R_1} \right\} \left\{ \left(\frac{\gamma_2}{\gamma_3}\right)^{R_2} \left(\frac{\gamma_3 - \gamma_2}{\gamma_3}\right)^{R_3 - R_2} \right\} ..... \left\{ \left(\frac{\gamma_{k-1}}{\gamma_k}\right)^{R_{k-1}} \left(\frac{\gamma_k - \gamma_{k-1}}{\gamma_k}\right)^{R_k - R_{k-1}} \right\}$$

These factors are respectively the probability given R_2 that the first two cells divide in the ratio $R_1 : R_2$ - R_1 ; the probability given R_3 that the proportion in cell 3 relative to cells 1 and 2 combined is R_3 - $R_2 : R_2$ and so on.

(5) The log likelihood function is,

$$l = n\left[\{Z_1\phi_1 - Z_2 g(\phi_1)\} + \{Z_2\phi_2 - Z_3 g(\phi_1)\} + ... + \{Z_{k-1}\phi_{k-1} - Z_k g(\phi_{k-1})\}\right]$$

where,

$$\phi_j = log\left\{\frac{\gamma_j}{\gamma_{j+1} - \gamma_j}\right\}$$

and

$$g\left(\phi\right) = log\left\{\frac{\gamma_{j+1}}{\gamma_{j+1} - \gamma_j}\right\}$$

(6) The iterative optimization method is used to update β until convergence to the maximum likelihood estimate.