FOML Assignment 1

Vishal Vijay Devadiga (CS21BTECH11061)

Abhay Kumar (BM21BTECH11001)

Question 2

Read Regression Models for Ordinal Data by Peter McCullagh

Part A

Provide a brief summary of the paper. Explain how the likelihood and odds ratio for ordinal regression is different from multi-class classification. How is it different from the regression problems?

Solution

Summary of the Paper   The paper discusses the problem of regression for ordinal data and how these models are multivariate extensions of generalized linear models.

The paper also dervies the likelihood function for ordinal regression and discusses existence and uniqueness of maximum likelihood estimates.

Let k be the number of classes with π_i being the probability of the ith class.

$$Link(\gamma_j) = \theta_j + \beta^T x$$

where:

- γ_j is the cumulative probability of the jth class.
- θ_j is the "cut point" for the jth class.
- β is the vector of coefficients.
- x is the vector of input features.
- $Link$ is the logit for proportional odds model and complementary log-log for proportional hazards model.

For the proportional odds model:

$$Link(\gamma_j) = \log\left[\frac{\gamma_j}{1-\gamma_j}\right]$$

For the proportional hazards model:

$$Link(\gamma_j) = \log\left[-\log(1-\gamma_j)\right]$$

Difference between Ordinal Regression and Multi-class Classification   In multi-class classification, the classes are mutually exclusive. However, in ordinal regression, the classes are ordered.

Likelihood function for ordinal regression uses the cumulative probabilities of the classes, whereas likelihood function for multi-class classification uses the probabilities of the classes.

Odds ratio for ordinal regression is the change in odds of the observed class falling in a higher category with a unit change in the predictor. Odds ratio for multi-class classification is not defined. Instead of odds ratio, we use the probability to compare the classes.

Difference between Ordinal Regression and Regression   In regression, the output is a continuous variable and is used to predict a continuous value.

In ordinal regression, the output is a discrete ordered variable and is used to predict a discrete ordered value.

Part B

Explain and derive the parameter estimation technique for the ordinal likelihood described in the paper.

Solution

Let k be the number of classes with π_i being the probability of the ith class. Let γ_j be the cumulative probability of the jth class.

Likelihood function for ordinal regression is given by:

$$L = \prod_{i=1}^{k} \pi_i^{n_i}$$

$$\pi_j = \gamma_j - \gamma_{j-1}$$

$$L = \prod_{i=1}^{k} \left[\gamma_i - \gamma_{i-1} \right]^{n_i}$$

$$\implies \log L = \sum_{i=1}^{k} n_i \log[\gamma_i - \gamma_{i-1}]$$

$$\implies \log L = n_1 \log \gamma_1 + \sum_{i=2}^{k} n_i \log[\gamma_i - \gamma_{i-1}]$$

Let $R_i = \sum_{j=1}^{i} n_j$

$$\implies \log L = R_1 \log \gamma_1 + \sum_{i=2}^{k} (R_i - R_{i-1}) \log[\gamma_i - \gamma_{i-1}]$$

$$\implies \log L = R_1 \log \gamma_1 + \sum_{i=2}^{k} (R_i - R_{i-1}) \log[\gamma_i - \gamma_{i-1}] + \sum_{i=2}^{k} -R_{i-1} \log \gamma_i - (R_i - R_{i-1} \log \gamma_i + R_i \log \gamma_i)$$

$$\implies \log L = \sum_{i=1}^{k-1} \left(R_i \log \frac{\gamma_i}{\gamma_{i+1}} + (R_{i+1} - R_i) \log \left[\frac{\gamma_{i+1} - \gamma_i}{\gamma_{i+1}} \right] \right)$$

$$\implies \log L = \sum_{i=1}^{k-1} \left(R_i \log \frac{\gamma_i}{\gamma_{i+1} - \gamma_i} + R_{i+1} \log \frac{\gamma_{i+1} - \gamma_i}{\gamma_{i+1}} \right)$$

Let $\phi_j = \log \dfrac{\gamma_j}{\gamma_{j+1} - \gamma_j}$

Let $g(\phi_j) = log(1 + \exp(\phi_j)) = \log \dfrac{\gamma_{j+1}}{\gamma_{j+1} - \gamma_j}$

Let $Z_i = \dfrac{R_i}{n}$

$$\boxed{\log L = n \sum_{i=1}^{k-1} Z_i \phi_i - Z_{i+1} g(\phi_i)}$$

To estimate the parameters, we need to maximize log likelihood.

Part C

Consider the WINE data set.

Develop an ordinal regression model in Python to predict the rating of the wine based on the input features.

Also, propose an appropriate evaluation metric and compare the result with a linear regression model.

Solution

The data set contains 11 input features and 1 output feature. The output feature is the quality of the wine, which is an integer between 3-9. The input features are: fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol. I have added a new column called type, which is 0 for red wine and 1 for white wine. Output feature is quality.

I have used the following models:

- `LogisticAT` from `mord` package for ordinal regression.
- `LinearRegression` from `sklearn` package for linear regression.

To evaluate the models, I have used the following metrics:

- Mean absolute error (MAE)
- R2 score
- Accuracy score
- Precision
- Recall
- F1 Score

All results are given in the `ipynb` file.