Name: Mayuresh Dindorkar
Roll No: CS23MTECH14007
Name: Sanyam Kaul
Roll No: CS23MTECH11011

PAGE No.
DATE / /

Q2

Ans 2

a) → This paper, written by Peter McCullagh, talks about building models on ordinal data by taking into account the ordinal nature of the data. This is very useful as it eliminates the need of assuming cardinality over data (by assigning some weights etc).

There are 2 data modes developed in the paper which are briefly discussed below:-

→ Proportional Odds Model:-

The model takes the ordinance of data as categories and then specifies odds ratios of different categories based on covariate values. A generalized emperical logit transform is also discussed in the paper which is used for parameter estimation.

Suppose that there are k ordered categories of the response and their corresponding probabilities are $p_1(x)$, $p_2(x)$ ---, $p_k(x)$ when covariates have value x. For a given category j, let $P(Y \leq j)$ denote the commulative probability of observing an outcome less than or equal to j. Then in the proportional odds model, we have:-

PAGE No.

DATE / /

$$P(Y \leq j|X) = P(X'\beta + \epsilon \leq \alpha_j)$$

$$= P(\epsilon \leq \alpha_j - X'\beta)$$

Here,

Y is the ordinal response variable
X is the vector of covariates
β is a vector of coefficients
ϵ is a random error term
α_j are the threshold parameters.

The proportional odds model has a key assumption that the odds ratio for any two individuals with covariates x_1 and x_2 remain constant across all categories:

$$\frac{P(Y \leq j|x_1)}{P(Y > j|x_1)} = \frac{P(Y \leq j|x_2)}{P(Y > j|x_2)}$$

→ Proportional Hazard Model :-

This model analyses the survival data. The model specifies a Hazard function, denoted by $h(t,X)$, which gives the instantaneous failure rate at time t for an individual with covariates X. The hazard function is defined as:-

$$h(t,X) = h_0(t) \cdot \exp(X'\beta)$$

PAGE No.
DATE / /

Here,
$h_0(t)$ is the baseline hazard function. This specifies the hazard when all covariates are 0.
X is the vector of covariates.
β is the vector of coefficients

Then, the model defines a survivor function which gives the probability of surviving beyond time t for an individual with covariate X. It is related to hazard function as follows:-

$$S(t, X) = \exp\left(-\int_0^t h(u, X)\, du\right)$$

$$= \exp\left(-\int_0^t h_0(u) \cdot \exp(X'\beta)\, du\right)$$

The key assumption of the Proportional Hazard Model is that the ratio of hazard functions (hazard ratio) is constant over time and across different levels of covariates:

$$\frac{h(t, x_1)}{h(t, x_2)} = \frac{\exp(x_1'\beta)}{\exp(x_2'\beta)} = \exp((x_1 - x_2)'\beta)$$

→ Both these models are shown to be the multivariate ~~exp~~ extensions of the Generalized linear models. Extensions to non-linear models are also discussed in the paper.

PAGE No.
DATE / /

→ The likelihood and odds ratio in proportional odds model are calculated for ordinal data and not the stocastic data. They are designed to capture the nature of ordinal response and how they are influenced by predicted variables.
In multi-class classification we predict classes without caring about any inherent order/ ranking among them.

→ It is different from ordinary regression in the following ways:-

1) In ordinal regression the data is categorised among different categories which can have meaningful progression among them (eg low, medium, high). In ordinary regression the response variable is continuous which means that it can take any value within a given range.

2) Ordinal regression aims to model the relationship between predictors and ordinal response. It then aims at estimating commulative probabilities of falling into each/ lower categories. This is unlike traditional regression which estimates a quantitative outcome.

PAGE No. DATE / /

b) → Deriving parameter estimation technique for proportional odds model:-

We have defined in part ⓐ that for proportional odds model:-

$$P(Y \le j|x) = P(x'\beta + \epsilon \le \alpha_j)$$

The likelihood function over a sample of n observations will then be given by:-

$$L(\beta) = \prod_{i=1}^{n} \prod_{j=1}^{k-1} P(Y=j|x_i)^{I(Y_i=j)}$$

∴ The log likelihood will be:-

$$\boxed{l(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{k-1} I(Y_i=j) \log P(Y=j|x_i)}$$

To find MLE we differentiate $l(\beta)$ with respect to β and set it equal to 0.

$$\boxed{\frac{\partial l}{\partial \beta} = \sum_{i=1}^{n} \sum_{j=1}^{k-1} I(Y_i=j) \frac{\partial}{\partial \beta} \log P(Y=j|x_i)}$$

From proportional odds model, we have:-

$$P(Y=j|x_i) = P(\alpha_{j-1} - x_i'\beta < \epsilon < \alpha_j - x_i'\beta)$$

Taking log of this we get:-

PAGE No.

DATE / /

$$\log P(Y=j|x_i) = \log\left(P(\epsilon > \alpha_{j-1} - x_i'\beta) - P(\epsilon > \alpha_j - x_i'\beta)\right)$$

$$= \log\left(S(\alpha_{j-1} - x_i'\beta) - S(\alpha_j - x_i'\beta)\right)$$

Here, $S(t) = 1 - F(t)$ is the survival function

$$\therefore \boxed{\frac{\partial l}{\partial \beta} = -\left(S(\alpha_{j-1} - x_i'\beta) - S(\alpha_j - x_i'\beta)\right) \cdot x_i}$$

On equating this to 0 we solve for MLE.

→ Deriving MLE for proportional hazard model:-

From (a) we know that:-

$$h(t,x) = h_0(t) \cdot \exp(x'\beta)$$, where $h(t,x)$ is hazard function

$$\therefore \boxed{L(\beta) = \prod_{i=1}^{n} [h(t_i, x_i)]^{\delta_i} \cdot S(t_i, x_i)}$$

$$l(\beta) = \sum_{i=1}^{n} \delta_i \log[h(t_i, x_i)] + \sum_{i=1}^{n} \log[S(t_i, x_i)]$$

Now, in order to learn β we partial derivate $l(\beta)$ wrt β:-

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^{n} \delta_i \cdot \frac{1}{h(t_i, x_i)} \cdot \frac{\partial h(t_i, x_i)}{\partial \beta} - \sum_{i=1}^{n} \frac{1}{S(t_i, x_i)} \cdot \frac{\partial S(t_i, x_i)}{\partial \beta}$$

Here, partial derivative of hazard function is:-

$$\boxed{\frac{\partial h(t_i, x_i)}{\partial \beta} = h_0(t) \cdot \exp(x'\beta) \cdot x}$$

Partial derivative of survivor function:-

$$\boxed{\frac{\partial S(t_i, x_i)}{\partial \beta} = -S(t,x) \cdot \exp(x'\beta) \cdot x}$$

We substitute this back and equate to 0 to find MLE