Cross Validated

Mini Project 1

Nirmal Agarwal(190557)
Priyanshu Yadav (190652)

For $i = 1, ..., n$, let $x_i = (1, x_{i2}, ..., x_{ip})$ be the vector of covariates for the i^{th} observation and $\beta \in R_p$ be the corresponding vector of regression coefficients. Suppose response y_i is a realization of Y_i with

$$Y_i \sim Bern(\phi(x_i^T\beta)),$$

where $\phi(.)$ is the CDF of a standard normal distribution.

Question 1

Write an algorithm for obtaining the maximum likelihood estimator of with all mathematical details.

Solution

$$Pr(Y_i = 1) = \alpha_i = \phi(x_i^T\beta)$$

$$\alpha_i = \frac{1}{\sqrt{2\pi}} \int_{-\inf}^{x_i^T\beta} e^{\frac{-u^2}{2}} du$$

Differentiating α

$$\nabla\alpha_i = \frac{x_i^T}{\sqrt{2\pi}} e^{\frac{-(x_i^T\beta)^2}{2}}$$

$$L(\beta|Y) = \prod_{i=1}^{n} (\alpha_i)^{y_i}(1-\alpha_i)^(1-y_i)$$

$$l(\beta) = \sum_{i=1}^{n} y_i log(\alpha_i) + \sum_{i=1}^{n} (1-y_i) log(1-\alpha_i)$$

Differentiating

$$\nabla l(\beta) = \sum_{i=1}^{n} \frac{y_i(\nabla\alpha_i)}{\alpha_i} + \sum_{i=1}^{n} \frac{(1-y_i)(-\nabla\alpha_i)}{1-\alpha_i}$$

$$\nabla l(\beta) = \sum_{i=1}^{n} \frac{y_i(\nabla\alpha_i) - \alpha_i y_i(\nabla\alpha_i) - \alpha_i(\nabla\alpha_i) + \alpha_i y_i(\nabla\alpha_i)}{\alpha_i(1-\alpha_i)}$$

$$\nabla l(\beta) = \sum_{i=1}^{n} \frac{y_i(\nabla\alpha_i) - \alpha_i(\nabla\alpha_i)}{\alpha_i(1-\alpha_i)}$$

$$\nabla l(\beta) = \sum_{i=1}^{n} \frac{(y_i-\alpha_i)(\nabla\alpha_i)}{\alpha_i(1-\alpha_i)}$$

$$\nabla l(\beta) = 0$$

We use gradient descent to solve for β.

Question 2

On April 10, 1912, the RMS Titanic sank after colliding with an iceberg on its maiden voyage. The accident killed 1502 of the 2224 passengers and crew on board. A subset of the data of the survival of passengers and crew can be loaded in R using titanic <- read.csv("titanic.csv"). Note: You must have the dataset saved in the same folder as R/Python script to load the dataset. The dataset has the response: whether a passenger survived the tragedy (Survived). Additional information includes the sex of the passenger (1 if male, 0 otherwise)(Sexmale),age (Age), the number of siblings/spouse aboard (SibSp), the number of parents/children aboard (Parch) and the passengers fare (Fare) (in British pounds). The data-set contains 712 observations. Obtain the MLE estimates of the corresponding β coefficients using the model described above.

Solution

The corresponding β are:
β=[0.39927561828259867, -1.3416911416628843, 0.33549291620798843, -0.09768899754299433, 0.003458524667663765, 1.2296688206058102]

Question 3

Jack Dawson was 20 years old when the tragedy happened. He boarded the ship with no family or spouse and paid 7.5 British pounds for his ticket. Rose Bukater was 19 years old on the tragic day. She boarded the ship with her fiance (treat this as a spouse) and her mother. She paid 512 British pounds for her ticket. What are the estimated probabilities of survival for Jack Dawson and Rose Bukater from your solution in part (b)?

Solution

Probability of Jack = 0.19993062711360204
Probability of Rose = 0.9465465336883306