TDS10 Final Project

Aleksandrov Boyan, 320961, b.aleksandrov@studenti.luiss.it

Nishtelkova Maria

Last compiled on: 25 November, 2025

Abstract

IF you wish, you may add here a short abstract of 100 words max.

Introduction

"adsadsds"

Dataset Description

(Write about heart.csv here)

Data Preparation

```
heartData <- read.csv("heart.csv")
head(heartData)
```

```
##   age    sex     place              cp trestbps chol   fbs      restecg
## 1  63   Male Cleveland   typical angina      145  233  TRUE lv hypertrophy
## 2  67   Male Cleveland    asymptomatic      160  286 FALSE lv hypertrophy
## 3  67   Male Cleveland    asymptomatic      120  229 FALSE lv hypertrophy
## 4  37   Male Cleveland     non-anginal      130  250 FALSE         normal
## 5  41 Female Cleveland atypical angina      130  204 FALSE lv hypertrophy
## 6  56   Male Cleveland atypical angina      120  236 FALSE         normal
##   thalch exang oldpeak       slope ca         thal hdc
## 1    150 FALSE     2.3 downsloping  0 fixed defect   0
## 2    108  TRUE     1.5        flat  3       normal   2
```

```
## 3    129  TRUE    2.6        flat  2 reversable defect   1
## 4    187 FALSE    3.5 downsloping  0              normal   0
## 5    172 FALSE    1.4   upsloping  0              normal   0
## 6    178 FALSE    0.8   upsloping  0              normal   0
```

read.csv()

Exploratory Data Analysis

Multinomial Logistic Regression — Theory

Question(1.1):

We are using multinomial Logistic Regression because the response variable can take more than 2 categories. For these categories there is a separate set of coefficients and we choose one as the baseline. The coefficients describe how the predictors(age, sex, chol etc.) affect the probability of belonging to each outcome category.

In this regression the response variable Y can take K-number of different categories. We have to pick one of the categories to be the baseline - category 0, for every other category - k = 1,2...,K-1.

The model shows the probability of an observation belonging to category K using the multinomial logistic regression function:

$$P(Y = k \mid X) = \frac{\exp(\beta_{0k} + \beta_k^T X)}{1 + \sum_{j=1}^{K-1} \exp(\beta_{0j} + \beta_j^T X)}$$

The probability of the baseline category is:

$$P(Y = 0 \mid X) = \frac{1}{1 + \sum_{j=1}^{K-1} \exp(\beta_{0j} + \beta_j^T X)}$$

Multinomial Logistic Regression

(Fit model + interpretation)

Model Evaluation

(Cross-validation)

Model Improvement

(Stepwise model / alternative model)

Binary Logistic Regression

(Create hdc01 + logistic model)

Model Comparison

(Compare multinomial vs binary)

Conclusion

(Brief summary)