

Applied Machine Learning and Predictive Modelling I: Modelling Stroke Data

Authors: Larissa Eisele, Fabian Lüthard, Yves Maillard

Module: Applied Machine Learning and Predictive Modelling I

Submitted on 10th of June, 2022

SUPERVISOR: MATTEO TANADINI AND DANIEL MEISTER

Table of Contents

1	Introduction	1
2	Importing Data	1
3	Methodology	1
4	Linear Model	1
5	Generalised Linear Model with family set to Poisson	1
6	Generalised Linear Model with family set to Binomial	1
7	Generalised Additive Model	1
8	Neural Network	1
9	Support Vector Machine	1
10	solve an optimisation problem	1

1 Introduction

2 Importing Data

```
stroke_data <- read_csv('./data/healthcare-dataset-stroke-data.csv')

## Rows: 5110 Columns: 12
## -- Column specification -----
## Delimiter: ","
## chr (6): gender, ever_married, work_type, Residence_type, bmi, smoking_status
## dbl (6): id, age, hypertension, heart_disease, avg_glucose_level, stroke
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
stroke_data

## # A tibble: 5,110 x 12
##       id gender  age hypertension heart_disease ever_married work_type
##   <dbl> <chr>  <dbl>         <dbl>         <dbl> <chr>         <chr>
## 1  9046 Male    67             0             1 Yes         Private
## 2  51676 Female  61             0             0 Yes         Self-employed
## 3  31112 Male    80             0             1 Yes         Private
## 4  60182 Female  49             0             0 Yes         Private
## 5   1665 Female  79             1             0 Yes         Self-employed
## 6  56669 Male    81             0             0 Yes         Private
## 7  53882 Male    74             1             1 Yes         Private
## 8  10434 Female  69             0             0 No          Private
## 9  27419 Female  59             0             0 Yes         Private
## 10 60491 Female  78             0             0 Yes         Private
## # ... with 5,100 more rows, and 5 more variables: Residence_type <chr>,
## #   avg_glucose_level <dbl>, bmi <chr>, smoking_status <chr>, stroke <dbl>
```

3 Methodology

4 Linear Model

5 Generalised Linear Model with family set to Poisson

6 Generalised Linear Model with family set to Binomial

7 Generalised Additive Model

8 Neural Network

9 Support Vector Machine

10 solve an optimisation problem