University of Oxford

MSc in Statistical Science

Final thesis

Missing data imputation for Haemorrhagic shock prediction

Author:
Antoine Ogier

Supervisor:
Pr. Julie Josse
(École polytechnique)
Pr. Geoff Nicholls
(University of Oxford)

September 2018

Abstract

Lorem ipsum dolot sit amet nunc cui Brexit.

Acknowledgements

Lorem ipsum dolot sit amet

Contents

Introduction

Chapter 1

Goal and data

1.1 The problem of haemorrhagic shock

Prediction is very hard, as described in [1]

1.2 The Traumabase data

1.3 Exploratory data analysis

1.3.1 Variables

1.3.2 Missing data

Chapter 2

Imputation methods

2.1 Main types of imputation

2.1.1 Joint parametric specification

2.1.2 Fully conditional specification: the MICE algorithm

2.1.3 Low-rank approximation for imputation

2.1.4 ML-based

2.2 Multiple imputation

2.2.1 Principle

2.2.2 Rubin's rule and prediction aggregation

2.3 Normality hypothesis: transforming the data

Chapter 3

Methodology: imputation and the validation split

3.1 Empirical risk minimization (ERM): classical context

3.2 The problem of current imputation methods

3.2.1 Imputation seen as an ERM

3.2.2 Unsuitability of current methods

3.3 Possible solutions

3.3.1 Using current implementations

3.3.2 A new variant: Multivariate Normal Mode with reserved data

3.3.3 Comparison on simulated data

Chapter 4

Error sources and best imputation: the case of linear regression with missing data

4.1 Problem set-up

4.1.1 Notations

4.1.2 Objective

4.2 Partial resolution

4.2.1 General loss

4.2.2 When the validation set is fully observed

Strong consistency of the least square estimator [2]

4.2.3 When the data is large and the training data is fully observed

4.3 Consequences

4.3.1 Theoretical implications for our data

4.3.2 Verification with simulated data

Chapter 5

Analysis: imputing the Traumabase data for prediction

5.1 Criteria for evaluation

5.2 Single imputation

5.3 Multiple imputation

Chapter 6

Results

Conclusion

Bibliography

[1] Matthew J Pommerening, Goodman, et al. Clinical gestalt and the prediction of massive transfusion after trauma. *Injury*, 46(5):807–813, 2015.

[2] TW Anderson, John B Taylor, et al. Strong consistency of least squares estimates in normal linear regression. *The Annals of Statistics*, 4(4): 788–790, 1976.