The School of Mathematics

My Incredible Thesis

by

My Name

Dissertation Presented for the Degree of
MSc in Statistics with Data Science

July 2021

Supervised by
Dr Very Important and Dr Strangelove

Executive Summary

Here comes your executive summary ...

Acknowledgments

Here come your acknowledgments ...

University of Edinburgh – Own Work Declaration

This sheet must be filled in, signed and dated - your work will not be marked unless this is done.

Name: . . . .
Matriculation Number: . . . .
Title of work: . . . .

I confirm that all this work is my own except where indicated, and that I have:

- Clearly referenced/listed all sources as appropriate
- Referenced and put in inverted commas all quoted text (from books, web, etc)
- Given the sources of all pictures, data etc. that are not my own
- Not made any use of the report(s) or essay(s) of any other student(s) either past or present
- Not sought or used the help of any external professional academic agencies for the work
- Acknowledged in appropriate places any help that I have received from others (e.g. fellow students, technicians, statisticians, external sources)
- Complied with any other plagiarism criteria specified in the Course handbook

I understand that any false claim for this work will be penalised in accordance with the University regulations (`https://teaching.maths.ed.ac.uk/main/msc-students/msc-programmes/statistics/data-science/assessment/academic-misconduct`).

Signature . . . .

Date . . . .

Contents

List of Tables

List of Figures

1 Introduction

In the introduction, I explain what do I want to know, and why. Here I will write a very good, precise and brief introduction. Particularly Section 3 is good!

2 Data Preprocessing

In this part, I applied exploratory data analysis (EDA) on specific columns in easySHARE dataset, including the visualization of data. Further, I selected appropriate variables to create risk factors of dementia, following the easySHARE data guide [2] and the work of Livingston et al [4], and extracted a subset of data relevant for the subsequent modelling. Finally, feature engineering was implemented on the subset for modelling steps.

2.1 Exploratory Data Analysis and Variable Selection

2.1.1 Cognitive Score

EasySAHRE dataset does not record diagnosis of dementia (e.g. Alzheimer's disease) in all waves. Instead, it contains the following indices to describe the cognitive functions of respondents:

- `recall_1`: the number of words recalled in the first trial of the word recall task, ranging from 0 to 10.
- `recall_2`: the number of words recalled in the delayed word recall task, ranging from 0 to 10.
- `orienti`: orientation of date, month, year and day of week, ranging from 0 (good) to 4 (bad).
- `numeracy_1`: information on the Mathematical performance, ranging from 1 (bad) to 5 (good).
- `numeracy_2`: information on the second test on Mathematical performance, ranging from 1 to 5.

Therefore, I considered creating a composite cognitive score based on the indices above as a proxy for dementia severity and using it as the response variable in subsequent modelling. Before any combination, I visualized the distributions of these columns using histograms:

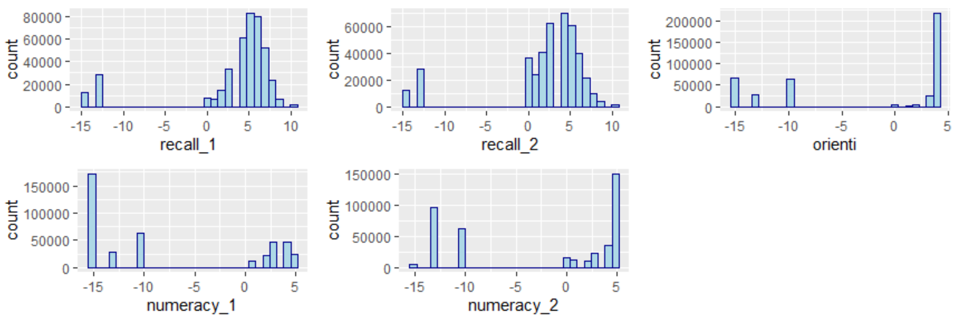

Figure 1: Distributions of cognitive function indices

3 Methods

In the following, I explain what did I do and how.

I should really cut this short, but BlaBlaBlaBla BlaBlaBlaBlaBla Bla Bla BlaBlaBla Bla Bla BlaBlaBla Bla BlaBla BlaBla Bla BlaBlaBlaBla Bla BlaBla Bla Bla Bla BlaBla BlaBlaBlaBla BlaBlaBlaBlaBla Bla Bla BlaBlaBla Bla. Bla BlaBlaBla Bla BlaBla BlaBla Bla BlaBlaBlaBla Bla BlaBla Bla Bla Bla BlaBla BlaBlaBlaBla. BlaBlaBlaBlaBla Bla Bla BlaBlaBla Bla Bla BlaBlaBla Bla BlaBla BlaBla Bla BlaBlaBlaBla Bla BlaBla Bla Bla Bla BlaBla BlaBlaBlaBla BlaBlaBlaBlaBla Bla Bla BlaBlaBla Bla Bla BlaBlaBla Bla BlaBla BlaBla Bla BlaBlaBlaBla Bla BlaBla Bla Bla Bla BlaBla BlaBlaBlaBla BlaBlaBlaBlaBla Bla Bla BlaBlaBla Bla Bla BlaBlaBla Bla BlaBla BlaBla Bla. BlaBlaBlaBla Bla BlaBla Bla Bla Bla BlaBla BlaBlaBlaBla BlaBlaBlaBlaBla Bla Bla BlaBlaBla Bla Bla BlaBlaBla Bla BlaBla BlaBla Bla BlaBlaBlaBla Bla BlaBla Bla Bla Bla BlaBla.

Note that I start a new paragraph when I have an empty line like this. BlaBlaBlaBla BlaBlaBlaBlaBla Bla Bla BlaBlaBla Bla Bla BlaBlaBla Bla BlaBla BlaBla Bla BlaBlaBlaBla Bla BlaBla Bla Bla Bla. BlaBla BlaBlaBlaBla BlaBlaBlaBlaBla Bla Bla BlaBlaBla Bla Bla BlaBlaBla Bla BlaBla BlaBla Bla BlaBlaBlaBla Bla BlaBla Bla Bla Bla BlaBla BlaBlaBlaBla BlaBlaBlaBlaBla Bla Bla BlaBlaBla Bla Bla BlaBlaBla. Bla BlaBla BlaBla Bla BlaBlaBlaBla Bla BlaBla Bla Bla Bla BlaBla BlaBlaBlaBla BlaBlaBlaBlaBla Bla Bla BlaBlaBla Bla Bla BlaBlaBla Bla. BlaBla BlaBla Bla BlaBlaBlaBla Bla BlaBla Bla Bla Bla BlaBla BlaBlaBlaBla BlaBlaBlaBlaBla Bla Bla BlaBlaBla Bla Bla BlaBlaBla Bla BlaBla BlaBla Bla BlaBlaBlaBla Bla BlaBla Bla Bla Bla BlaBla.
But I can also end a line with a double backslash.

3.1 Models

Models are *very* helpful because.

- They're good.
- They're helpful.

3.2 Techniques

Techniques even better because.

1. They're magnificent.
2. If they work.

4 Results

I this section, I explain what did I discover.

Now it's getting very technical ... I will cite [3, 1]. I will also show my incredible α, β and γ mathematics and do some other fancy stuff.

4.1 Formulae

For example look at this

$$\min \sum_{s\in\mathcal{S}} Pr_s \left[\sum_{t=1}^{T} \left(\sum_{g\in\mathcal{G}} \left(\alpha_{gts} C_g^0 + p_{gts} C_g^1 + (p_{gts})^2 C_g^2 \right) + \sum_{g\in\mathcal{C}} \gamma_{gts} C_g^s \right) \right], \tag{4.1}$$

and you will see that it has a little number on the side so that I can refer to it as equation (4.1). Now if I do this

$$\begin{aligned} \sum_{i=1}^{n} k_i &= 20 \quad (4.2) \\ \sum_{j=20}^{m} \delta_i &\geq \eta \end{aligned}$$

I can align two formulae and control which one has a number on the side. It is (4.2). I can also do something like this

$$Y_l = \begin{bmatrix} \left(y_s + i\frac{b_c}{2}\right)\frac{1}{\tau^2} & -y_s\frac{1}{\tau e^{-i\theta^s}} \\ -y_s\frac{1}{\tau e^{i\theta^s}} & y_s + i\frac{b_c}{2} \end{bmatrix},$$

and it won't have a number on the side. Now if I have to do some huge mathematics I'd better structure it a little and include linebreaks etc. so that it fits on one page.

$$\begin{aligned} p_l^f &= G_{l11}\left(2v_{F(l)}\bar{v}_{F(l)} - \bar{v}_{F(l)}^2\right) \quad (4.3) \\ &+ \bar{v}_{F(l)}\bar{v}_{T(l)}\left[B_{l12}\sin\left(\bar{\delta}_{F(l)} - \bar{\delta}_{T(l)}\right) + G_{l12}\cos\left(\bar{\delta}_{F(l)} - \bar{\delta}_{T(l)}\right)\right] \\ &+ \begin{bmatrix} \bar{v}_{T(l)}\left[B_{l12}\sin\left(\bar{\delta}_{F(l)} - \bar{\delta}_{T(l)}\right) + G_{l12}\cos\left(\bar{\delta}_{F(l)} - \bar{\delta}_{T(l)}\right)\right] \\ \bar{v}_{F(l)}\left[B_{l12}\sin\left(\bar{\delta}_{F(l)} - \bar{\delta}_{T(l)}\right) + G_{l12}\cos\left(\bar{\delta}_{F(l)} - \bar{\delta}_{T(l)}\right)\right] \\ \bar{v}_{F(l)}\bar{v}_{T(l)}\left[B_{l12}\cos\left(\bar{\delta}_{F(l)} - \bar{\delta}_{T(l)}\right) - G_{l12}\sin\left(\bar{\delta}_{F(l)} - \bar{\delta}_{T(l)}\right)\right] \\ \bar{v}_{F(l)}\bar{v}_{T(l)}\left[-B_{l12}\cos\left(\bar{\delta}_{F(l)} - \bar{\delta}_{T(l)}\right) + G_{l12}\sin\left(\bar{\delta}_{F(l)} - \bar{\delta}_{T(l)}\right)\right] \end{bmatrix} \cdot \begin{bmatrix} v_{F(l)} - \bar{v}_{F(l)} \\ v_{T(l)} - \bar{v}_{T(l)} \\ \delta_{F(l)} - \bar{\delta}_{F(l)} \\ \delta_{T(l)} - \bar{\delta}_{T(l)} \end{bmatrix}, \end{aligned}$$

This is a lot of fun!

4.2 Important Things

Finally we should have a nice picture like this one. However, I won't forget that figures and table are environments which float around in my document. So LaTeX will place them wherever it thinks they fit well with the surrounding text. I can try to change that with a float specifier, e.g. [!ht]. Now I

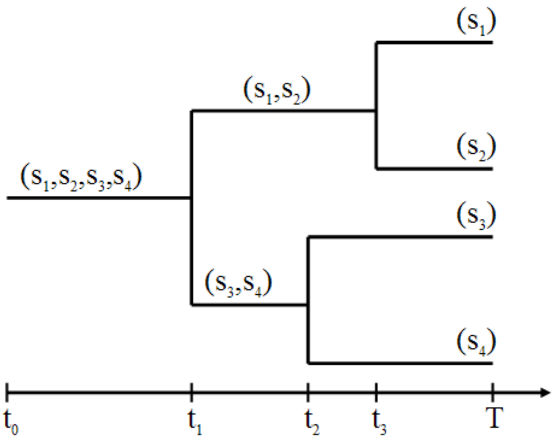

Figure 2: Look at this scenario tree with funny times t_1 and scenarios s_1 etc.

want to use one of my own environments. I want to define something.

Definition 4.1 *I define*

$$\Gamma_\eta := \sum_{i=1}^{n}\sum_{j=i}^{n}\xi(i,j)$$

I definitely need some good tables, so I do this. I should really refer to Table 1.

Case	Generators	Therm. Units	Lines	Peak load: [MW]	[MVar]
6 bus	3 at 3 buses	2	11	210	210
9 bus	3 at 3 buses	3	9	315	115
24 bus	33 at 11 buses	26	38	2850	580
30 bus	6 at 6 buses	5	41	189.2	107.2
39 bus	10 at 10 buses	7	46	6254.2	1387.1
57 bus	7 at 7 buses	7	80	1250.8	336.4

Table 1: Something that doesn't make sense.

4.3 And now something else

Let:

$$\begin{aligned}
\Omega_0 &= \{(x,y,z,f):\ \text{satisfying } (9)-(19)\},\\
\Omega_1 &= \{(x,y,z,f):\ \text{satisfying } (9),(11)-(20)\},\\
\overline{\Omega}_0 &= \{\mathbf{0}\leq(x,y,z,f)\leq\mathbf{1}:\ \text{satisfying } (9)-(18)\},\\
\overline{\Omega}_1 &= \{\mathbf{0}\leq(x,y,z,f)\leq\mathbf{1}:\ \text{satisfying } (9),(11)-(18),(20)\}.
\end{aligned}$$

where **0** and **1** are vectors of appropriate dimensions with 0's and 1's, respectively. Next we see that both Ω_0 and Ω_1 give equivalent formulations for the A-MSSP. In particular, the following statements hold:

Proposition 1 $\Omega_0 \subseteq \Omega_1$.

Proof. Let us suppose there exists $(x, y, z, f) \in \Omega_1$ such that $(x, y, z, f) \notin \Omega_0$. Then, there exist indices $i \in I$ and $t \in \{0, \dots, |T| - s_i\}$ with $x_i^t > 0.5 \left(\sum_{h=1}^{s_i} x_i^{t+h} + 1 \right)$. By definition, $x_i^t = 1$ and $x_i^{t+h} = 0$ for all $h \in \{1, \dots, s_i\}$. By (11) and (12), $\sum_{h=1}^{s_i} f_i^{th} = 1$, so $f_i^{th'} = 1$ for some $h' \in \{1, \dots, s_i\}$. But then,

$$0 = x_i^{t+h'} = \sum_{h=\max\{1, t+h'-(|T|-s_i)\}}^{\min\{s_i, t+h'\}} f_i^{t+h'-h,h} \geq f_i^{th'} = 1,$$

as $h' \in [\max\{1, t + h' - (|T| - s_i)\}, \min\{s_i, t + h'\}]$. □

This immediately gives us

Corollary 1 *AS is a valid formulation for the A-MSSP.*

Next we compare the Linear Programming (LP) relaxations of the two formulations.

Proposition 2 $\overline{\Omega}_1 \subseteq \overline{\Omega}_0$.

Proof. Homework □

5 Conclusions

I have no idea how to conclude, so I don't write much. But the stuff that follows is important.

References

[1] N. Gröwe-Kuska and W. Römisch. *Stochastic unit commitment in hydro-thermal power production planning.* Preprints aus dem Institut für Mathematik. Humboldt-Universität zu Berlin, Institut für Mathematik, 2001.

[2] N. Gröwe-Kuska and W. Römisch. *Stochastic unit commitment in hydro-thermal power production planning.* Preprints aus dem Institut für Mathematik. Humboldt-Universität zu Berlin, Institut für Mathematik, 2001.

[3] T. Shiina and J. R. Birge. Stochastic unit commitment problem. *International Transactions in Operational Research*, 11(1):19–32, 2004.

[4] T. Shiina and J. R. Birge. Stochastic unit commitment problem. *International Transactions in Operational Research*, 11(1):19–32, 2004.

Appendices

A An Appendix

Some stuff.

B Another Appendix

Some other stuff.