Vysoká škola ekonomická v Praze

Fakulta informatiky a statistiky

Variační autoenkodér a úlohy pozorování v latentním prostoru

BAKALÁŘSKÁ PRÁCE

Studijní program: Aplikovaná informatika

Studijní obor: Aplikovaná informatika

Autor: Tomáš Faltejsek

Vedoucí práce: Ing. Ondřej Vadinský, Ph.D.

Konzultant práce: full consultant's name (incl. degrees)

Praha, květen 2023

Poděkování

Thanks.

Abstract

Jedním z předních rysů lidské inteligence je intuice a schopnost představovat si nové objekty. Variační autoenkodér je inovací na poli pravděpodobnostních modelů, umožňující architekturu modelů schopných syntézy zcela nových dat s využitím pozorování atributů v latentním prostoru. Teoretická charakteristika a možnosti využití variačního autoenkodéru jsou předmětem této bakalářské práce.

Keywords

keyword, important term, another topic, and another one

Obsah

Seznam obrázků

Note: Add a list of figures if the number of figures in the thesis text exceeds 20. A list of diagrams is applicable only if the author distinguishes between a figure and a diagram. The list of diagrams is included if the number of diagrams exceeds 20. This thesis template does not distinguish between a figure and a diagram.

Seznam tabulek

Note: Add a list of tables if the number of tables used in the thesis exceeds 20.

Seznam použitých zkratek

BCC Blind Carbon Copy

CC Carbon Copy

CERT Computer Emergency Response Team

CSS Cascading Styleheets

DOI Digital Object Identifier

HTML Hypertext Markup Language

REST Representational State Transfer

SOAP Simple Object Access Protocol

URI Uniform Resource Identifier

URL Uniform Resource Locator

XML eXtended Markup Language

Note: Add a list of abbreviations if the number of abbreviations used in the thesis exceeds 20 and the abbreviations used are not common.

Introduction

Introduction is a compulsory part of the bachelor's / diploma thesis. The introduction is an introduction to the topic. It elaborates the chosen topic, briefly puts it into context (there may also be a description of the motivation to write the work) and answers the question why the topic was chosen. It puts the topic into context and justifies its necessity and the topicality of the solution. It contains an explicit goal of the work. The text of the thesis goal is identical with the text that is given in the bachelor's thesis assignment, ie with the text that is given in the InSIS system and which is also given in the Abstract section.

Part of the introduction is also a brief introduction to the process of processing the work (a separate part of the actual text of the work is devoted to the method of processing). The introduction may also include a description of the motivation to write the work.

The introduction to the diploma thesis must be more elaborate - this is stated in more detail in the Requirements of the diploma thesis within the Intranet for FIS students.

Here are some sample chapters that recommend how a bachelor's / master's thesis should be set. They primarily describe the use of the LaTeX template, but general advice will also serve users of other systems well.

1. Tables, pictures, programs, formulas

The use of tables and graphs/figures in technical text has some common rules and some specific ones. We do not present tables and graphs/figures directly in the text, but we place them either on separate pages or in a reserved place at the top or bottom of regular pages. LaTeX will take care of the placement of floating graphs and tables automatically.

Graphs/figures and tables are numbered and equipped with a legend. The legend should describe the content of the graph or table in such detail that the reader understands them without a thorough study of the text of the work.

There must be a numerical reference to the table and graph/figure in the text (a dynamic cross-reference mechanism, which is part of LaTeX, can be strongly recommended). At the appropriate place in the text, we then summarize the most important conclusions that can be drawn from a table or graph. The text should be legible and understandable even without looking at the tables and graphs, and the tables and graphs should be understandable even without reading the text in detail.

We refer to tables and graphs, if possible, indirectly during the normal flow of text; instead *„Table 1.1 shows that men are on average about* 9,9 kg *heavier than women"* we'd rather write *„Men are* 9,9 kg *heavier then women (see tab. 1.1)".*

1.1 Tables

Tabulka 1.1: Maximum plausible estimates in model M.

| Effect	Estimate	Standard. deviation[a]	P-value
Abs. member	−10,01	1,01	—
Gender (male)	9,89	5,98	0,098
Height (cm)	0,78	0,12	< 0,001

Pozn:[a] Standard error of estimation by Monte Carlo method.

The following tips specifically apply to **tables**:

- Never copy tables from statistical software to a thesis. Typically, statistical software also includes more information in tables than necessary.
- Avoid vertical lines. Thicker horizontal lines separate the table from the surrounding text, including the legend, weaker horizontal lines separate the column headers from the table body, and the individual parts of the table from each other. In LaTeX, the `booktabs` package implements this form of tables. If we want to significantly separate some columns from others, we insert a larger space between them

- Keep the type, format and sense of the field content in a single column. It is not advised to enter, e.g., average and percent in the same column.
- Avoid repeating the same field content too many times. E.g., if the column Variance shows the value of 0.5 in the first ten lines and 1.5 in the following ten lines, cancel the column. Find a different solution. E.g., one can divide the table to two. Alternatively, one can enter descriptive lines that inform of a variable value repeating in the following table section. E. g. „*Variance* = 0,5 " and below „*Variance* = 1,5 ").
- All numbers shall have the same number of valid digits. Numbers in a table shall be aligned to the decimal point.
- A table sometimes requires the use of abbreviations that do not occur elsewhere. Such abbreviations may be explained in the legend or notes below the table. Notes below the table may also be used for an explanation of the sense of some columns or values.

1.2 Figures
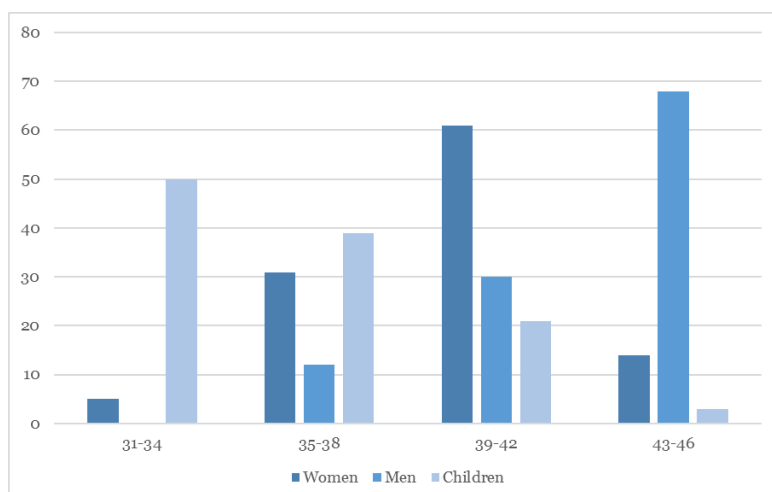
Obrázek 1.1: Frequency of shoe size in the population of men, women and children (CZSO data, Author's calculation)

There are several general tips for figures and diagrams.

- A figure/diagram should be created in the same size as used in the thesis. Decreasing a large diagram leads to having unreadable labels. Increasing a small diagram leads to poor graphical quality.
- The diagram axis shall be properly labelled in the thesis language. Missing punctuation is tolerable. If a diagram deals with, e.g., weight and height, the labels shall say *Height [cm]* and *Weight [kg]*. If the graph includes the function *h(x)*, the axes get a label of *x* and *h(x)*. Each axis shall bear a clearly defined scale.
- If a two-dimensional diagram marks many points, the author should make sure that they do not get mixed. If the number of points is too high, the author should decrease the size of the symbols that refer to them or select a lower number of points to mark

in the diagram. Diagrams with thousands of marked points cause problems mainly in electronic documents by increasing the file size.

- If the thesis is to be printed in black and white, the author should avoid using colours. Lines should be distinguished by line type (full, dotted, dashed. . .). Sections should be distinguished by distinct shades of grey or hatching. The sense of individual line types or hatched sections shall be explained either in the diagram textual legend or in a graphic legend integrated into the diagram.
- Avoid bitmap figures with a low resolution, especially JPEGs. Compression artifacts do not look good on paper.

1.3 Source codes

Algorithms, programme excerpts and descriptions of programme interactions shall be distinguished from other text sections. One option is to use package `listings`, which defines a simple `code` environment in the `makra.tex` file. It can be used to create, for example, the following examples.

```
> mean(x)
[1]  158.90
> object$mean
[1]  158.90
```

However, the `listings` package and its `lstlisting` environment offer an almost inexhaustible number of configuration parameters, eg for highlighting the syntax of programming languages (several tens), line numbering, etc. Examples:

- https://en.wikibooks.org/wiki/LaTeX/Source_Code_Listings
- https://www.overleaf.com/learn/latex/Code_listing#Using_listings_to_highlight_code

1.4 Typesetting of mathematics

We type the variables in italics (TeX does this in math mode itself, but don't forget that in the surrounding text and also turn on math mode). We place function names upright. For example: $\mathrm{var}(X) = \mathsf{E}\, X^2 - \left(\mathsf{E}\, X\right)^2$.

Fractions inside a paragraph (e. g. $\frac{5}{7}$ or $\frac{x+y}{2}$) they can be too cramped, so it's better to bet simple fractions with a slash: $5/7$, $(x + y)/2$.

The possibilities of LaTeX for typesetting mathematics are rich, but they may not be sufficient in some specific situations. Therefore, American Mathematical Society (AMS) packages

can be recommended for use. The `makra.tex` file loads the `amsmath`, `amsfonts` and `amsthm` packages by default. To penetrate their possibilities, the following will serve:

- Math Extension with AMS-L^AT_EX – `http://ptgmedia.pearsoncmg.com/images/032117 3856/samplechapter/kopkach15.pdf`
- `https://www.overleaf.com/learn/latex/Aligning_equations_with_amsmath`
- Math Mode – `http://tex.loria.fr/general/Voss-Mathmode.pdf`
- More Math into LaTeX – `http://tug.ctan.org/info/Math_into_LaTeX-4/Short_Co urse.pdf`

Example of a numbered formula:

$$\mathbf{b} = (\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{y} \tag{1.1}$$

Example of unnumbered formulas with functions and indexes:

$$d_{ij} = \max_{k=1,2,\ldots,n} \{d_{ik} + d_{kj}\},$$
$$x_{1,2} = b \pm \sqrt{\ln y}.$$

An example of a formula as part of one paragraph is given on the example of supplier capacities in a mathematical model of a traffic problem, which we take into account using constraints:

$$\sum_{j=1}^{n} x_{ij} \leq a_i, \qquad i = 1,2,\ldots,m , \tag{1.2}$$

where expression a_i represents capacity of i-th supplier.

When deriving a formula by incremental modification, the individual steps are usually listed on separate lines (`align*` environment from the `amsmath` package):

$$f(x) = (x + a)(x + b) =$$
$$= x^2 + bx + ax + ab =$$
$$= x^2 + (a + b)x + ab$$

Example of column adjustment (`eqnarray*`):

$$\sum_{i=1}^{n} x_{ij} = 1, \qquad j = 1,2,\ldots,n,$$
$$\sum_{j=1}^{n} x_{ij} = 1, \qquad i = 1,2,\ldots,n,$$
$$u_i + 1 - M(1 - x_{ij}) \leq u_j, \qquad i = 2,3,\ldots,n, \quad j = 1,2,\ldots,n,$$
$$u_i \geq 0, \qquad i = 1,2,\ldots,n,$$
$$x_{ij} \in \{0,1\} \qquad i = 1,2,\ldots,n, \quad j = 1,2,\ldots,n,$$

2. Work with literature

The template assumes the use of a bibliographic database for greater flexibility. The use of a bibliographic database is not a necessary condition, it is possible to make do with the standard environment `thebibliography`. In this case, however, it is necessary to perform interventions on some files, as described below.

2.1 Use of bibliographic database

1. **Change the database name**
   The template assumes a database stored in a file `bibliography.bib`. If the database is named differently, then it is necessary in the file `makra.tex` change the value of the `\bibliography` command parameter.
2. **Change citation style**
   By default, citations in the text are given in numerical variant. You can easily switch to using a combination of last name and year by changing the file `makra.tex`, where the comment character is swapped in the parameters for the package `biblatex`.

2.2 Use of the environment `thebibliography`

1. In the file `makra.tex` at the beginning delete these lines:

```
%%% Nastavení pro použití samostatné bibliografické databáze.
%%% Settings for using a separate bibliographic database.
\usepackage[
    backend=biber
%  ,style=iso-authoryear
   ,style=iso-numeric
   ,sortlocale=cs_CZ
   ,alldates=iso
   ,bibencoding=UTF8
   ,maxnames=2
   ,maxbibnames=99
   %,block=ragged
]{biblatex}
\let\cite\parencite
\renewcommand*{\multinamedelim}{, \addspace}
\renewcommand*{\finalnamedelim}{\addspace a \addspace}

\bibliography{bibliography}
```

2. In the file `bibliography.tex` delete the line `\printbibliography` and remove the comment flag in the next section containing the environment `thebibliography.`

2.3 How cite in the text

$$
\begin{aligned}
&\texttt{\textbackslash cite\{Cermak2018\}} &&\longrightarrow (1) \\
&\texttt{\textbackslash cite\{Hladik2018,Jasek2018\}} &&\longrightarrow (2, 3) \\
&\texttt{\textbackslash cite[chap. 3]\{Pecakova2018\}} &&\longrightarrow (4, \text{kap. } 3)
\end{aligned}
$$

3. PDF/A format

Electronic form of final work must be submitted in PDF/A format level 1a or 2u. They are PDF profiles that determine which PDF properties are allowed to use to make the documents suitable for long-term archiving and further automatic processing. Next we will deal with level 2u, which we bet on LaTeX.

The most important requirements of PDF/A-2u include:

- All fonts must be built into the document. They are not allowed links to external fonts.
- Fonts must contain a ToUnicode table that defines the conversion from encoding characters used inside a Unicode font. This makes it possible from the document reliably extract text.
- The document must contain metadata in XMP format and, if colored, then also the formal specification of color space.

This template uses the tt pdfx package, which LaTeX can set up to meet the requirements of PDF/A. Metadata in XMP is generated automatically by information in the file tt thesis.xmpdata (you can refer to the generated file see in tt pdfa.xmpi).

The correctness of PDF/A can be checked using an online validator: `https://www.pdf-online.com/osa/validate.aspx/`.

If the file is not valid, common causes include less use common fonts (which are inserted only in bitmap format and/or without unicode tables) and embedding images in PDF, which are standard in themselves PDF/A do not meet.

This is likely to be the case for images created by many different programs. In this case, you can try to convert the image to PDF/A using GhostScript, for example, as follows:

```
gs -q -dNOPAUSE -dBATCH
   -sDEVICE=pdfwrite -dPDFSETTINGS=/prepress
   -sOutputFile=vystup.pdf vstup.pdf
```

Conlusion

The conclusion is a mandatory part of the bachelor's / diploma thesis. It contains a summary of the work and comments on the degree of fulfillment of the goal, which was set in the work, or summarizes the answers to the questions that were asked in the introduction.

The conclusion to the diploma thesis must be more elaborate - this is stated in more detail in the Requirements of the diploma thesis within the Intranet for FIS students.

The conclusion is perceived as a chapter, which begins on a separate page and is called the conclusion. The name Conclusion is not numbered. The text of the conclusion itself is divided into paragraphs.

Bibliografie

1. ČERMÁK, Radim a SMUTNÝ, Zdeněk. A Framework for Cultural Localization of Websites and for Improving Their Commercial Utilization. In: *Global Observations of the Influence of Culture on Consumer Buying Behavior.* Hershey: IGI Global, 2018, s. 206–232. ISBN 978-1-5225-2727-5. Dostupné z DOI: `10.4018/978-1-5225-2727-5.ch013`.

2. HLADÍK, Milan a ČERNÝ, Michal. The Shape of the Optimal Value of a Fuzzy Linear Programming Problem. In: *Fuzzy Logic in Intelligent System Design.* Cham: Springer, 2018, s. 281–286. Advances in Intelligent Systems and Computing 648. ISBN 978-3-319-67136-9. Dostupné z DOI: `10.1007/978-3-319-67137-6_31`.

3. JAŠEK, Pavel, VRANÁ, Lenka, ŠPERKOVÁ, Lucie, SMUTNÝ, Zdeněk a KOBULSKÝ, Marek. Modeling and Application of Customer Lifetime Value in Online Retail. *Informatics.* 2018, roč. 5, č. 1. Dostupné také z: `http://www.mdpi.com/2227-9709/5/1/2/pdf`.

4. PECÁKOVÁ, Iva. *Statistika v terénních průzkumech.* Praha: Professional Publishing, 2018. ISBN 978-80-88260-10-3.

Přílohy

A. Full form

B. Source codes of computational procedures