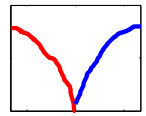
Predictive Analysis of Customer Credit Data

Yousuf Hasan Siddiqui
Orhan Ipek
Shikhar Sirivastava

Statistical Programming Language
Humboldt–Universität zu Berlin

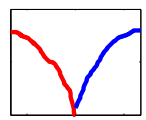
Outline

1. Introduction ✓
2. Motivation
3. Data Set
4. Descriptive Statistics
5. Encoding Data
6. Predictive Modelling
7. Efficency Tests
8. Results
9. Conclusion



Introduction

- `Beamer` is the latest package to create slides with LaTeX
- Slides need to be compiled to PDF, not DVI/Postscript
- Remember: PDFLaTeX accepts PNG, JPEG and PDF not EPS/PS
- If you *need* Postscript, RTFM

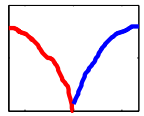


Motivation

- The LvB Beamer Style is defined via *beamerdefs.sty*, *colordef.sty* and *lvblisting.sty*, which must always be provided in the source folder.
- All operators are to be defined by `\operatorname{}`. Note the difference:

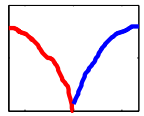
  Var defined by operatorname
  Var not defined by operatorname

- Remember to start and end the displaymath environment by `\[` and `\]` and not `$$`.



Data Set

- German Credit Data
- Mixture of Quantitative and Qualitative Attributes
- Observations: 1000 , Variables: 20
- Preprocessing Data:
  - Point 1
  - Point 2
- Some other points

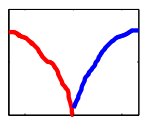


Equations

- Equations covering several lines may be written in the *align* environment instead of the older *eqnarray* environment. Only this way it can be ensured, that the colour of the equation and of the according equation numbering match.
- `align*` omits the equation numbering, as does `\notag`.

```
\begin{align}
4x + 8 &= (3-2)^2\\
4x &= -7 \notag \\
x &= -\frac{7}{4}
\end{align}
```

$$4x + 8 = (3 - 2)^2 \tag{1}$$
$$4x = -7$$
$$x = -\frac{7}{4} \tag{2}$$

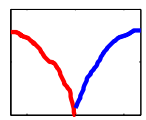


Tables

Title	Title
2.13	1.45
3.14	6.85

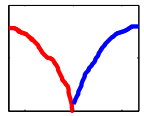
Table 1: Include a short, but meaningful caption.

- Follow the Cambridge University Press Style.
- Not more than 2 decimal digits in a column.
- Tables and their captions are to be written in black.



Tables

```
\begin{table}
\begin{center}
\begin{tabular}{cc}
\hline\hline
Title & Title\\
\hline
2.13 & 1.45 \\
3.14 & 6.85 \\
\hline\hline
\end{tabular}
\caption{Include a short, but meaningful caption.}
\end{center}
\end{table}
```

Figures

```
1 \begin{figure}[htb]
2     \begin{center}
3     \includegraphics[
        scale=0.2]{
        Figures/vola}
4     \caption{Include a
         short, but
        meaningful
        caption.}
5     \end{center}
6 \end{figure}
```

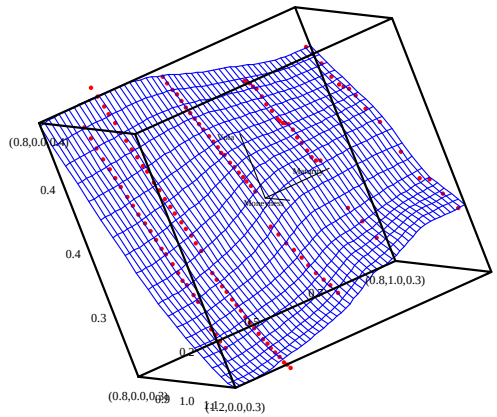
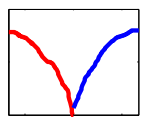


Figure 1: Include a short, but meaningful caption.

The caption is, as in tables, to be written in black and please provide any legend in the caption and not in the graph itself.



Examples

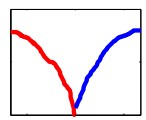
To create an example, use the color `isegreen` and the following structure:

```
\color{isegreen}
\textbf{Example:} Example title

\smallskip
Here you can state your example, which may also include calculations.
\color{black}
```

Example: Example title

Here you can state your example, which may also include calculations.



Subtitles

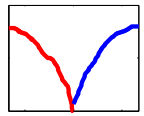
Subtitles are to be highlighted via bold text and followed by a small skip afterwards (no colon):

```
1 \textbf{Subtitle}
2
3 \smallskip
4 Here you can state the
    content according to
    the subtitle.
```

Subtitle

Here you can state the content according to the subtitle.

This may also be applied to state proofs, theorems etc.

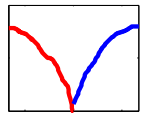


Brackets

- Use the bracket sequence $[\{(a+b=c)\}]$
- Conventional bracket rules represent an exemption of this rule. For example:

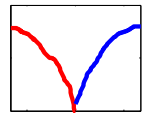
$$Y \sim \mathrm{N}(\mu(X), \sigma(X))$$

- Let LaTeX take care about the correct size by preceding the bracket by `\left` and `\right`.



Rules to write nice slides

- Use `\section{}` and `\subsection{}` to structure your presentation. The section will appear in the upper right corner of your slides.
- You can set up hyperlinks via `\label{LINKNAME}` (reference point) and `\ref{LINKNAME}` (reference).
- Use, if necessary, `\displaystyle` to force LaTeX to display fractions in big font size
- Remember
  - 6-8 lines per slide
  - 8 words per line



- The numbering of any enumeration should match the colour of the corresponding text (preset colour: black). Modifications may be made through the *itemize* environment:

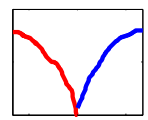
  `\item[\color{isegreen}1.]`

  Itemize items are predefined (blue) and excluded from this rule.
- Use `^{\top}` to write the symbol of transpose, it produces

$$x^{\top} y$$

- Use `\ldots` to write the symbol for three dots, it produces

$$x \in \{1, \ldots, n\}$$

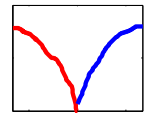


- The commands `\widehat{}` and `\widetilde{}` for a hat or a tilde are to be preferred over the the smaller `\hat` respectively `\tilde` commands:

$$\widehat{Y} \text{ vs. } \hat{Y}$$
$$\widetilde{Y} \text{ vs. } \tilde{Y}$$

- The norm is to be written via `\|`. It produces $\|K\|$
- The $\mathcal{O}$ and $\mathcal{o}$ for convergence may be written via `\mathcal{O}` and `\mbox{\scriptsize $\mathcal{O}$}`.
- The operator for exponential terms with Euler's e as the base is defined by `\exp`:

$$\exp(1) \approx 2.718282$$



- Use `\stackrel{\mathcal{L}}{\rightarrow}` to write the symbol for convergence in distribution and denote the normal distribution by `\operatorname{N}`, this produces

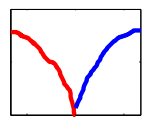
$$X \stackrel{\mathcal{L}}{\rightarrow} \mathrm{N}(0, \sigma^2)$$

- Use `\operatorname{P}` to write the symbol for probability, it produces

$$\mathrm{P}(X = x) = \frac{\exp(-\lambda)\lambda^x}{x!}$$

- Use `\stackrel{\operatorname{as.}}{\sim}` to write the symbol for asymptotic distribution, it produces

$$X \stackrel{\text{as.}}{\sim} \chi^2$$



- Use command `\stackrel{\operatorname{def}}{=}` to write the symbol for definition, it produces

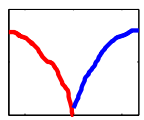
$$X \stackrel{\text{def}}{=} \frac{a}{b}$$

- Use commands `\Re` or `\Im` to write the symbols for the real or imaginary part, it produces

$$X = \Re\{Y\}, Y = \Im\{Z\}$$

- To write the symbols for the minimizing argument, use `\operatorname{arg}\,\underset{x}{\operatorname{min}}`, it produces

$$a = \arg\min_{x}\{f(x)\}$$



- Use `\operatorname{\mathbf{I}}` for the indicator function:

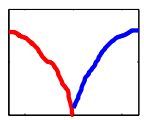
$$\mathbf{I}\{x < 1\}$$

- Use `\ln` or `\log` to write the symbols for natural logarithm or decimal logarithm, it produces

$$1 = \ln(\exp(1)), \quad 1 = \log(10)$$

- Use `\operatorname{E}` to write the symbol for expectation, it produces

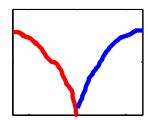
$$\mathrm{E}[X] = \mu$$



Using `listings` for source

Slides containing a listing also need [containsverbatim] as option.
For 'highlighting' of XploRe keywords see `listing.tex`.

```
1 library("metrics")
2 randomize(10178)
3 z=(uniform(n).>0.5)~(normal(n).<0.5)
```

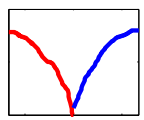



Piecewise Uncovering I

The following example uses $< 1 - 2 >$ commands to piecewise hide and uncover text. $< 1 - 2 >$ makes the first item appear only on slides 1 and 2, $< 2- >$ has the second item visible from slide 2 onwards.

- Itemize environments

(i) First Roman point.



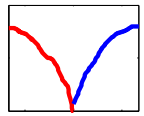
Piecewise Uncovering I

The following example uses $< 1 - 2 >$ commands to piecewise hide and uncover text. $< 1 - 2 >$ makes the first item appear only on slides 1 and 2, $< 2- >$ has the second item visible from slide 2 onwards.

- Itemize environments
- can be uncovered or hidden

(i) First Roman point.

(ii) Second Roman point, uncovered on second slide.

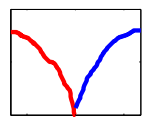


Piecewise Uncovering I

The following example uses $< 1 - 2 >$ commands to piecewise hide and uncover text. $< 1 - 2 >$ makes the first item appear only on slides 1 and 2, $< 2- >$ has the second item visible from slide 2 onwards.

- can be uncovered or hidden
- piecewise.

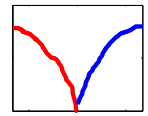
(i) First Roman point.
(ii) Second Roman point, uncovered on second slide.
(iii) Last Roman point.



Piecewise Uncovering II

There is an easier way using \item $<+->$

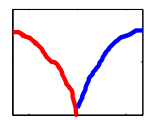
- Itemize environments



Piecewise Uncovering II

There is an easier way using \item < +− >

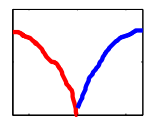
- Itemize environments
- can be uncovered or hidden



Piecewise Uncovering II

There is an easier way using \item $<+->$

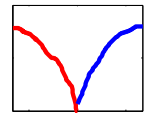
- Itemize environments
- can be uncovered or hidden
- piecewise.



Hiding text. . .

Text on the first slide.

Shown on all slides.

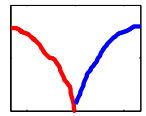


Hiding text. . .

Text on the first slide.
Shown on second and third slide.

- Still shown on 2nd and 3rd slide.

Shown on all slides.



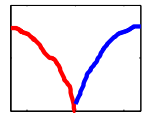
Hiding text. . .

Text on the first slide.
Shown on second and third slide.

- Still shown on 2nd and 3rd slide.

- Shown on slides 3 and 5.

Shown on all slides.

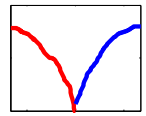


Hiding text. . .

Text on the first slide.

- Shown from slide 4 on.

Shown on all slides.

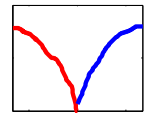


Hiding text. . .

Text on the first slide.

- Shown from slide 4 on.
- Shown on slides 3 and 5.

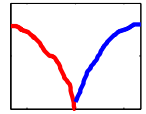
Shown on all slides.



Further Information

Further Information can be found in the LaTeX version of this document, where some more details are explained and important specifications are highlighted.

Suggestions to improve the style or the explanations are welcome!

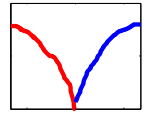


For Further Reading

Tobias Oetiker, Hubert Partl, Irene Hyna and Elisabeth Schlegl
The Not So Short Introduction to LaTeX2e
available on www.ctan.org, 2008

Scott Pakin
The Comprehensive LaTeXSymbol List
available on www.ctan.org, 2008

Frank Mittelbach and Michel Goossens
The LaTeX Companion – 2nd ed.
Addison-Wesley, 2004



For Further Reading

Mark Trettin and Jürgen Fenn
An essential guide to LaTeX2e usage
available on www.ctan.org, 2007

Wikipedia Wiki Books
LaTeX-Wörterbuch: InDeX
available on www.wikipedia.de

Till Tantau
User Guide to the Beamer Class, Version 3.07
available on www.sourceforge.net, 2007