Chaotic
Black Boxes

The risks and opportunites of deriving knowledge from deep machine learning models.

Brad Flaugher

September 21, 2022

We have now accumulated sufficient evidence to see that whatever language the central nervous system is using, it is characterized by less logical and arithmetical depth than what we are normally used to.

– John von Neumann

Children can learn to use computers in a masterful way, and ... learning to use computers can change the way they learn everything else.

– Seymour A. Papert

Preface

I am of the opinion that every LaTeX geek, at least once during his life, feels the need to create his or her own class: this is what happened to me and here is the result, which, however, should be seen as a work still in progress. Actually, this class is not completely original, but it is a blend of all the best ideas that I have found in a number of guides, tutorials, blogs and tex.stackexchange.com posts. In particular, the main ideas come from two sources:

- Ken Arroyo Ohori's Doctoral Thesis, which served, with the author's permission, as a backbone for the implementation of this class;
- The Tufte-Latex Class, which was a model for the style.

The first chapter of this book is introductory and covers the most essential features of the class. Next, there is a bunch of chapters devoted to all the commands and environments that you may use in writing a book; in particular, it will be explained how to add notes, figures and tables, and references. The second part deals with the page layout and design, as well as additional features like coloured boxes and theorem environments.

I started writing this class as an experiment, and as such it should be regarded. Since it has always been intended for my personal use, it may not be perfect but I find it quite satisfactory for the use I want to make of it. I share this work in the hope that someone might find here the inspiration for writing his or her own class.

Federico Marotta

Contents

List of Figures

List of Tables

List of Listings

History: the slow march away from algorithms

Ages of Understanding | 1

1.1 (1500-Today) Algorithms: Codified Human Understanding

AI is a shitty term

We tried a lot of things, teaching computers explicit grammar and explicit rules

IMO, this was not AI, this was codified human understanding.

In code, that understanding might look like this....

bradflaugher.com. [1] [1]

1.2 (1980-Today) Machine Learning: Data-Derived Insights

Hardware got amazing, we gave up teaching the way we teach ourselves and let the data do the work

We leveraged huge statistical models to regress our way to success

We used building blocks of regression and neurons to train huge models

These models are statistical and deterministic, but ultimately chaotic black boxes..

1: Snarky sidenote!

[1]: Andreu et al. (2021), *Humans won't be able to control a superintelligent AI, according*

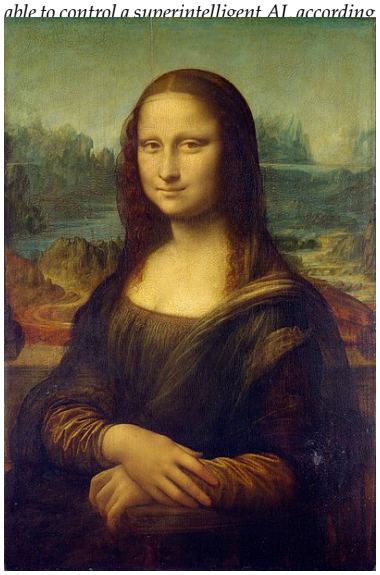

Figure 1.1: The Mona Lisa. https://commons.wikimedia.org/wiki/File:Mona_Lisa,_by_Leonardo_da_Vinci,_from_C2RMF_retouched.jpg

Deep Learning Concepts: layered statistical representations

How Models Read Data 2

2.1 Numerical Data

This is some text and a link to Hey if you want to site something on the side use[1]

2.2 Words

2.3 Sounds

2.4 Images

2.5 Video

2.6 Mixed Datasets

```
cd myproject
docker run tensorflow
#profit!
```

tex.stackexchange.org for help.

[1]: Andreu et al. (2021), *Humans won't be able to control a superintelligent AI, according to a study*

Learning Methods 3

3.1 Unsupervised

3.2 Supervised

3.3 Reinforcement

3.4 Domain Transfer

are you predicting the right thing? Are you really predicting how valuable the company is or just whether it'll be the next meme stock?

3.5 Notes on Ethics

Representation, "fixing the training set" [2], or the Impossibility of Fairness from a model.

[2]: Christian (2020), *The Alignment Problem: Machine Learning and Human Values*

```
cd myproject
docker run tensorflow
#profit!
```

`tex.stackexchange.org` for help.

The Chaotic Black Box: statistical inference with a billion or so parameters

Classifiers 4

4.1 Recommenders

4.2 Facial Recognition

4.3 Sentiment

4.4 Hate Speech

4.5 The Ethics of Classification

Online Advertising, Justice, Job Applications, Creditworthiness, Getting Insurance (Weapons of Math Destruction), Civic Life, /sideciteOneil2017 ; The Default Male, Invisible Women effects snow clearing schedules and drug discovery

```
cd myproject
docker run tensorflow
#profit!
```

tex.stackexchange.org for help.

Transformers 5

5.1 Style Transfer

5.2 Translation

5.3 Text Generation

GPT-3, BERT and Bloom

5.4 Image Generation and Stable Diffusion

Link some cool shit here, Draw Owl!

5.5 The Ethics of Transforming

Ensembles and Mathematical Chaos

6

6.1 Interacting Layers of Statistical Understanding

6.2 Useful Chaos

Appendix

A

Poetry Test

Let's say we want to build an ensemble model to analyze poetry, put a haiku into craiyon's online shit, then we categorize the resulting photo. [1]

[1]: Andreu et al. (2021), *Humans won't be able to control a superintelligent AI, according to a study*

Bibliography

Here are the references in citation order.

[1] Abraham Andreu and Qayyah Moynihan. 'Humans won't be able to control a superintelligent AI, according to a study'. In: *Business Insider* (Sept. 24, 2021). (Visited on 09/24/2021) (cited on pages 2, 4, 11).

[2] Brian Christian. *The Alignment Problem: Machine Learning and Human Values*. New York, NY, USA: W. W. Norton & Company, Oct. 2020 (cited on page 5).

Notation

The next list describes several symbols that will be later used within the body of the document.

c	Speed of light in a vacuum inertial frame
h	Planck constant

Alphabetical Index