

UNIVERSITAT POLITÈCNICA DE CATALUNYA

TREBALL FINAL DEL GRAU EN ENGINYERIA FÍSICA

Adaptive Computation Time

Author:

Daniel FOJO

Supervisor:

Dr. Xavier GIRÓ

Image Processing Group

Signal Theory and Communications Department

January, 2018

“Thanks to my solid academic training, today I can write hundreds of words on virtually any topic without possessing a shred of information, which is how I got a good job in journalism.”

Dave Barry

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Abstract

TelecomBCN

Signal Theory and Communications Department

Engineering Physics

Adaptive Computation Time

by Daniel FOJO

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too...

Acknowledgements

The acknowledgments and the people to thank go here, don't forget to include your project advisor...

Contents

Abstract	iii
Acknowledgements	v
1 Introduction	1
1.1 Welcome and Thank You	1
1.2 Learning L ^A T _E X	1
1.2.1 A (not so short) Introduction to L ^A T _E X	1
1.2.2 A Short Math Guide for L ^A T _E X	2
1.2.3 Common L ^A T _E X Math Symbols	2
1.2.4 L ^A T _E X on a Mac	2
1.3 Getting Started with this Template	2
1.3.1 About this Template	2
1.4 What this Template Includes	3
1.4.1 Folders	3
1.4.2 Files	3
1.5 Filling in Your Information in the <code>main.tex</code> File	4
1.6 The <code>main.tex</code> File Explained	5
1.7 Thesis Features and Conventions	6
1.7.1 Printing Format	6
1.7.2 Using US Letter Paper	6
1.7.3 References	6
A Note on bibtex	7
1.7.4 Tables	7
1.7.5 Figures	7
1.7.6 Typesetting mathematics	9
1.8 Sectioning and Subsectioning	9
1.9 In Closing	9
2 Introduction to Machine Learning	11
2.1 Machine Learning Basics	11
2.1.1 The task T	11
2.1.2 The Performance Measure, P	12
2.1.3 The Experience, E	12
2.2 Capacity, Overfitting and Underfitting	12
2.2.1 Regularization	13
2.2.2 The No Free Lunch Theorem	13
2.3 Gradient-Based Optimization	14
2.3.1 Gradient Descent	14
2.3.2 Stochastic Gradient Descent	14
2.3.3 Momentum	14

3	Introduction to Deep Learning	17
3.1	Main Section 1	17
3.1.1	Subsection 1	17
3.1.2	Subsection 2	17
3.2	Main Section 2	17
4	Adaptive Computation Time	19
4.1	Main Section 1	19
4.1.1	Subsection 1	19
4.1.2	Subsection 2	19
4.2	Main Section 2	19
5	Experiments	21
5.1	Addition	21
5.1.1	Subsection 1	21
5.1.2	Subsection 2	21
5.2	Parity	22
6	Conclusions	23
6.1	Main Section 1	23
6.1.1	Subsection 1	23
6.1.2	Subsection 2	23
6.2	Main Section 2	23
A	Frequently Asked Questions	25
A.1	How do I change the colors of links?	25
	Bibliography	27

List of Figures

1.1	An Electron	8
2.1	Here we have three examples of a regression. The training data was generated synthetically, by randomly sampling x values and choosing y deterministically by evaluating a quadratic function. (A) A linear function cannot capture the curvature and underfits the data. (B) A quadratic function fits to the data and generalizes well to unseen points. It does not suffer from overfitting or underfitting. (C) A polynomial of degree 9 overfits the data and does not generalize well to unseen points (Goodfellow, Bengio, and Courville, 2016).	13
2.2	Momentum	15

List of Tables

1.1	The effects of treatments X and Y on the four groups studied.	8
-----	---	---

List of Abbreviations

SGD	S tochastic G radient D escent
NN	N eural N etwork
RNN	R ecurrent N eural N etwork
CNN	C onvolutional N eural N etwork
LSTM	L ong S hort T erm M emory

For/Dedicated to/To my...

Chapter 1

Introduction

1.1 Welcome and Thank You

Welcome to this L^AT_EX Thesis Template, a beautiful and easy to use template for writing a thesis using the L^AT_EX typesetting system.

If you are writing a thesis (or will be in the future) and its subject is technical or mathematical (though it doesn't have to be), then creating it in L^AT_EX is highly recommended as a way to make sure you can just get down to the essential writing without having to worry over formatting or wasting time arguing with your word processor.

L^AT_EX is easily able to professionally typeset documents that run to hundreds or thousands of pages long. With simple mark-up commands, it automatically sets out the table of contents, margins, page headers and footers and keeps the formatting consistent and beautiful. One of its main strengths is the way it can easily typeset mathematics, even *heavy* mathematics. Even if those equations are the most horribly twisted and most difficult mathematical problems that can only be solved on a super-computer, you can at least count on L^AT_EX to make them look stunning.

1.2 Learning L^AT_EX

L^AT_EX is not a WYSIWYG (What You See is What You Get) program, unlike word processors such as Microsoft Word or Apple's Pages. Instead, a document written for L^AT_EX is actually a simple, plain text file that contains *no formatting*. You tell L^AT_EX how you want the formatting in the finished document by writing in simple commands amongst the text, for example, if I want to use *italic text for emphasis*, I write the `\emph{text}` command and put the text I want in italics in between the curly braces. This means that L^AT_EX is a "mark-up" language, very much like HTML.

1.2.1 A (not so short) Introduction to L^AT_EX

If you are new to L^AT_EX, there is a very good eBook – freely available online as a PDF file – called, "The Not So Short Introduction to L^AT_EX". The book's title is typically shortened to just *lshort*. You can download the latest version (as it is occasionally updated) from here: <http://www.ctan.org/tex-archive/info/lshort/english/lshort.pdf>

It is also available in several other languages. Find yours from the list on this page: <http://www.ctan.org/tex-archive/info/lshort/>

It is recommended to take a little time out to learn how to use L^AT_EX by creating several, small 'test' documents, or having a close look at several templates on:

<http://www.LaTeXTemplates.com>

Making the effort now means you're not stuck learning the system when what you *really* need to be doing is writing your thesis.

1.2.2 A Short Math Guide for L^AT_EX

If you are writing a technical or mathematical thesis, then you may want to read the document by the AMS (American Mathematical Society) called, “A Short Math Guide for L^AT_EX”. It can be found online here: <http://www.ams.org/tex/amslatex.html> under the “Additional Documentation” section towards the bottom of the page.

1.2.3 Common L^AT_EX Math Symbols

There are a multitude of mathematical symbols available for L^AT_EX and it would take a great effort to learn the commands for them all. The most common ones you are likely to use are shown on this page: <http://www.sunilpatel.co.uk/latex-type/latex-math-symbols/>

You can use this page as a reference or crib sheet, the symbols are rendered as large, high quality images so you can quickly find the L^AT_EX command for the symbol you need.

1.2.4 L^AT_EX on a Mac

The L^AT_EX distribution is available for many systems including Windows, Linux and Mac OS X. The package for OS X is called MacTeX and it contains all the applications you need – bundled together and pre-customized – for a fully working L^AT_EX environment and work flow.

MacTeX includes a custom dedicated L^AT_EX editor called TeXShop for writing your ‘.tex’ files and BibDesk: a program to manage your references and create your bibliography section just as easily as managing songs and creating playlists in iTunes.

1.3 Getting Started with this Template

If you are familiar with L^AT_EX, then you should explore the directory structure of the template and then proceed to place your own information into the *THESIS INFORMATION* block of the `main.tex` file. You can then modify the rest of this file to your unique specifications based on your degree/university. Section 1.5 on page 4 will help you do this. Make sure you also read section 1.7 about thesis conventions to get the most out of this template.

If you are new to L^AT_EX it is recommended that you carry on reading through the rest of the information in this document.

Before you begin using this template you should ensure that its style complies with the thesis style guidelines imposed by your institution. In most cases this template style and layout will be suitable. If it is not, it may only require a small change to bring the template in line with your institution’s recommendations. These modifications will need to be done on the `MastersDoctoralThesis.cls` file.

1.3.1 About this Template

This L^AT_EX Thesis Template is originally based and created around a L^AT_EX style file created by Steve R. Gunn from the University of Southampton (UK), department of Electronics and Computer Science. You can find his original thesis style file at his site, here: <http://www.ecs.soton.ac.uk/~srg/softwaretools/document/templates/>

Steve's `ecsthesis.cls` was then taken by Sunil Patel who modified it by creating a skeleton framework and folder structure to place the thesis files in. The resulting template can be found on Sunil's site here: <http://www.sunilpatel.co.uk/thesis-template>

Sunil's template was made available through <http://www.LaTeXTemplates.com> where it was modified many times based on user requests and questions. Version 2.0 and onwards of this template represents a major modification to Sunil's template and is, in fact, hardly recognisable. The work to make version 2.0 possible was carried out by **Vel** and Johannes Böttcher.

1.4 What this Template Includes

1.4.1 Folders

This template comes as a single zip file that expands out to several files and folders. The folder names are mostly self-explanatory:

Appendices – this is the folder where you put the appendices. Each appendix should go into its own separate `.tex` file. An example and template are included in the directory.

Chapters – this is the folder where you put the thesis chapters. A thesis usually has about six chapters, though there is no hard rule on this. Each chapter should go in its own separate `.tex` file and they can be split as:

- Chapter 1: Introduction to the thesis topic
- Chapter 2: Background information and theory
- Chapter 3: (Laboratory) experimental setup
- Chapter 4: Details of experiment 1
- Chapter 5: Details of experiment 2
- Chapter 6: Discussion of the experimental results
- Chapter 7: Conclusion and future directions

This chapter layout is specialised for the experimental sciences, your discipline may be different.

Figures – this folder contains all figures for the thesis. These are the final images that will go into the thesis document.

1.4.2 Files

Included are also several files, most of them are plain text and you can see their contents in a text editor. After initial compilation, you will see that more auxiliary files are created by \LaTeX or BibTeX and which you don't need to delete or worry about:

example.bib – this is an important file that contains all the bibliographic information and references that you will be citing in the thesis for use with BibTeX. You can write it manually, but there are reference manager programs available that will create and manage it for you. Bibliographies in \LaTeX are a large subject and you may need to read about BibTeX before starting with this. Many modern reference

managers will allow you to export your references in BibTeX format which greatly eases the amount of work you have to do.

MastersDoctoralThesis.cls – this is an important file. It is the class file that tells L^AT_EX how to format the thesis.

main.pdf – this is your beautifully typeset thesis (in the PDF file format) created by L^AT_EX. It is supplied in the PDF with the template and after you compile the template you should get an identical version.

main.tex – this is an important file. This is the file that you tell L^AT_EX to compile to produce your thesis as a PDF file. It contains the framework and constructs that tell L^AT_EX how to layout the thesis. It is heavily commented so you can read exactly what each line of code does and why it is there. After you put your own information into the *THESIS INFORMATION* block – you have now started your thesis!

Files that are *not* included, but are created by L^AT_EX as auxiliary files include:

main.aux – this is an auxiliary file generated by L^AT_EX, if it is deleted L^AT_EX simply regenerates it when you run the main **.tex** file.

main.bbl – this is an auxiliary file generated by BibTeX, if it is deleted, BibTeX simply regenerates it when you run the **main.aux** file. Whereas the **.bib** file contains all the references you have, this **.bbl** file contains the references you have actually cited in the thesis and is used to build the bibliography section of the thesis.

main.blg – this is an auxiliary file generated by BibTeX, if it is deleted BibTeX simply regenerates it when you run the main **.aux** file.

main.lof – this is an auxiliary file generated by L^AT_EX, if it is deleted L^AT_EX simply regenerates it when you run the main **.tex** file. It tells L^AT_EX how to build the *List of Figures* section.

main.log – this is an auxiliary file generated by L^AT_EX, if it is deleted L^AT_EX simply regenerates it when you run the main **.tex** file. It contains messages from L^AT_EX, if you receive errors and warnings from L^AT_EX, they will be in this **.log** file.

main.lot – this is an auxiliary file generated by L^AT_EX, if it is deleted L^AT_EX simply regenerates it when you run the main **.tex** file. It tells L^AT_EX how to build the *List of Tables* section.

main.out – this is an auxiliary file generated by L^AT_EX, if it is deleted L^AT_EX simply regenerates it when you run the main **.tex** file.

So from this long list, only the files with the **.bib**, **.cls** and **.tex** extensions are the most important ones. The other auxiliary files can be ignored or deleted as L^AT_EX and BibTeX will regenerate them.

1.5 Filling in Your Information in the main.tex File

You will need to personalise the thesis template and make it your own by filling in your own information. This is done by editing the **main.tex** file in a text editor or your favourite LaTeX environment.

Open the file and scroll down to the third large block titled *THESIS INFORMATION* where you can see the entries for *University Name*, *Department Name*, etc ...

Fill out the information about yourself, your group and institution. You can also insert web links, if you do, make sure you use the full URL, including the **http://** for this. If you don't want these to be linked, simply remove the **\href{url}{name}** and only leave the name.

When you have done this, save the file and recompile `main.tex`. All the information you filled in should now be in the PDF, complete with web links. You can now begin your thesis proper!

1.6 The `main.tex` File Explained

The `main.tex` file contains the structure of the thesis. There are plenty of written comments that explain what pages, sections and formatting the L^AT_EX code is creating. Each major document element is divided into commented blocks with titles in all capitals to make it obvious what the following bit of code is doing. Initially there seems to be a lot of L^AT_EX code, but this is all formatting, and it has all been taken care of so you don't have to do it.

Begin by checking that your information on the title page is correct. For the thesis declaration, your institution may insist on something different than the text given. If this is the case, just replace what you see with what is required in the *DECLARATION PAGE* block.

Then comes a page which contains a funny quote. You can put your own, or quote your favourite scientist, author, person, and so on. Make sure to put the name of the person who you took the quote from.

Following this is the abstract page which summarises your work in a condensed way and can almost be used as a standalone document to describe what you have done. The text you write will cause the heading to move up so don't worry about running out of space.

Next come the acknowledgements. On this page, write about all the people who you wish to thank (not forgetting parents, partners and your advisor/supervisor).

The contents pages, list of figures and tables are all taken care of for you and do not need to be manually created or edited. The next set of pages are more likely to be optional and can be deleted since they are for a more technical thesis: insert a list of abbreviations you have used in the thesis, then a list of the physical constants and numbers you refer to and finally, a list of mathematical symbols used in any formulae. Making the effort to fill these tables means the reader has a one-stop place to refer to instead of searching the internet and references to try and find out what you meant by certain abbreviations or symbols.

The list of symbols is split into the Roman and Greek alphabets. Whereas the abbreviations and symbols ought to be listed in alphabetical order (and this is *not* done automatically for you) the list of physical constants should be grouped into similar themes.

The next page contains a one line dedication. Who will you dedicate your thesis to?

Finally, there is the block where the chapters are included. Uncomment the lines (delete the % character) as you write the chapters. Each chapter should be written in its own file and put into the *Chapters* folder and named `Chapter1`, `Chapter2`, etc... Similarly for the appendices, uncomment the lines as you need them. Each appendix should go into its own file and placed in the *Appendices* folder.

After the preamble, chapters and appendices finally comes the bibliography. The bibliography style (called *authoryear*) is used for the bibliography and is a fully featured style that will even include links to where the referenced paper can be found online. Do not underestimate how grateful your reader will be to find that a reference to a paper is just a click away. Of course, this relies on you putting the URL information into the BibTeX file in the first place.

1.7 Thesis Features and Conventions

To get the best out of this template, there are a few conventions that you may want to follow.

One of the most important (and most difficult) things to keep track of in such a long document as a thesis is consistency. Using certain conventions and ways of doing things (such as using a Todo list) makes the job easier. Of course, all of these are optional and you can adopt your own method.

1.7.1 Printing Format

This thesis template is designed for double sided printing (i.e. content on the front and back of pages) as most theses are printed and bound this way. Switching to one sided printing is as simple as uncommenting the *oneside* option of the `documentclass` command at the top of the `main.tex` file. You may then wish to adjust the margins to suit specifications from your institution.

The headers for the pages contain the page number on the outer side (so it is easy to flick through to the page you want) and the chapter name on the inner side.

The text is set to 11 point by default with single line spacing, again, you can tune the text size and spacing should you want or need to using the options at the very start of `main.tex`. The spacing can be changed similarly by replacing the *singlespacing* with *onehalfspacing* or *doublespacing*.

1.7.2 Using US Letter Paper

The paper size used in the template is A4, which is the standard size in Europe. If you are using this thesis template elsewhere and particularly in the United States, then you may have to change the A4 paper size to the US Letter size. This can be done in the margins settings section in `main.tex`.

Due to the differences in the paper size, the resulting margins may be different to what you like or require (as it is common for institutions to dictate certain margin sizes). If this is the case, then the margin sizes can be tweaked by modifying the values in the same block as where you set the paper size. Now your document should be set up for US Letter paper size with suitable margins.

1.7.3 References

The `bibtex` package is used to format the bibliography and inserts references such as this one (**Reference1**). The options used in the `main.tex` file mean that the in-text citations of references are formatted with the author(s) listed with the date of the publication. Multiple references are separated by semicolons (e.g. (**Reference2**; **Reference1**)) and references with more than three authors only show the first author with *et al.* indicating there are more authors (e.g. (**Reference3**)). This is done automatically for you. To see how you use references, have a look at the `Chapter1.tex` source file. Many reference managers allow you to simply drag the reference into the document as you type.

Scientific references should come *before* the punctuation mark if there is one (such as a comma or period). The same goes for footnotes¹. You can change this but the most important thing is to keep the convention consistent throughout the thesis. Footnotes themselves should be full, descriptive sentences (beginning with a capital

¹Such as this footnote, here down at the bottom of the page.

letter and ending with a full stop). The APA6 states: “Footnote numbers should be superscripted, [...], following any punctuation mark except a dash.” The Chicago manual of style states: “A note number should be placed at the end of a sentence or clause. The number follows any punctuation mark except the dash, which it precedes. It follows a closing parenthesis.”

The bibliography is typeset with references listed in alphabetical order by the first author’s last name. This is similar to the APA referencing style. To see how L^AT_EX typesets the bibliography, have a look at the very end of this document (or just click on the reference number links in in-text citations).

A Note on bibtex

The bibtex backend used in the template by default does not correctly handle unicode character encoding (i.e. "international" characters). You may see a warning about this in the compilation log and, if your references contain unicode characters, they may not show up correctly or at all. The solution to this is to use the biber backend instead of the outdated bibtex backend. This is done by finding this in `main.tex`: `backend=bibtex` and changing it to `backend=biber`. You will then need to delete all auxiliary BibTeX files and navigate to the template directory in your terminal (command prompt). Once there, simply type `biber main` and biber will compile your bibliography. You can then compile `main.tex` as normal and your bibliography will be updated. An alternative is to set up your LaTeX editor to compile with biber instead of bibtex, see [here](#) for how to do this for various editors.

1.7.4 Tables

Tables are an important way of displaying your results, below is an example table which was generated with this code:

```
\begin{table}
\caption{The effects of treatments X and Y on the four groups studied.}
\label{tab:treatments}
\centering
\begin{tabular}{l l l}
\toprule
\thead{Groups} & \thead{Treatment X} & \thead{Treatment Y} \\
\midrule
1 & 0.2 & 0.8 \\
2 & 0.17 & 0.7 \\
3 & 0.24 & 0.75 \\
4 & 0.68 & 0.3 \\
\bottomrule
\end{tabular}
\end{table}
```

You can reference tables with `\ref{<label>}` where the label is defined within the table environment. See `Chapter1.tex` for an example of the label and citation (e.g. Table [1.1](#)).

1.7.5 Figures

There will hopefully be many figures in your thesis (that should be placed in the *Figures* folder). The way to insert figures into your thesis is to use a code template like this:

TABLE 1.1: The effects of treatments X and Y on the four groups studied.

Groups	Treatment X	Treatment Y
1	0.2	0.8
2	0.17	0.7
3	0.24	0.75
4	0.68	0.3

```

\begin{figure}
\centering
\includegraphics{Figures/Electron}
\decoRule
\caption[An Electron]{An electron (artist's impression).}
\label{fig:Electron}
\end{figure}

```

Also look in the source file. Putting this code into the source file produces the picture of the electron that you can see in the figure below.



FIGURE 1.1: An electron (artist's impression).

Sometimes figures don't always appear where you write them in the source. The placement depends on how much space there is on the page for the figure. Sometimes there is not enough room to fit a figure directly where it should go (in relation to the text) and so \LaTeX puts it at the top of the next page. Positioning figures is the job of \LaTeX and so you should only worry about making them look good!

Figures usually should have captions just in case you need to refer to them (such as in Figure 1.1). The `\caption` command contains two parts, the first part, inside the square brackets is the title that will appear in the *List of Figures*, and so should be short. The second part in the curly brackets should contain the longer and more descriptive caption text.

The `\decoRule` command is optional and simply puts an aesthetic horizontal line below the image. If you do this for one image, do it for all of them.

L^AT_EX is capable of using images in pdf, jpg and png format.

1.7.6 Typesetting mathematics

If your thesis is going to contain heavy mathematical content, be sure that L^AT_EX will make it look beautiful, even though it won't be able to solve the equations for you.

The “Not So Short Introduction to L^AT_EX” (available on CTAN) should tell you everything you need to know for most cases of typesetting mathematics. If you need more information, a much more thorough mathematical guide is available from the AMS called, “A Short Math Guide to L^AT_EX” and can be downloaded from: <ftp://ftp.ams.org/pub/tex/doc/amsmath/short-math-guide.pdf>

There are many different L^AT_EX symbols to remember, luckily you can find the most common symbols in [The Comprehensive L^AT_EX Symbol List](#).

You can write an equation, which is automatically given an equation number by L^AT_EX like this:

```
\begin{equation}
E = mc^2
\label{eqn:Einstein}
\end{equation}
```

This will produce Einstein's famous energy-matter equivalence equation:

$$E = mc^2 \tag{1.1}$$

All equations you write (which are not in the middle of paragraph text) are automatically given equation numbers by L^AT_EX. If you don't want a particular equation numbered, use the unnumbered form:

```
\[ a^2=4 \]
```

1.8 Sectioning and Subsectioning

You should break your thesis up into nice, bite-sized sections and subsections. L^AT_EX automatically builds a table of Contents by looking at all the `\chapter{}`, `\section{}` and `\subsection{}` commands you write in the source.

The Table of Contents should only list the sections to three (3) levels. A `\chapter{}` is level zero (0). A `\section{}` is level one (1) and so a `\subsection{}` is level two (2). In your thesis it is likely that you will even use a `\subsubsection{}`, which is level three (3). The depth to which the Table of Contents is formatted is set within `MastersDoctoralThesis.cls`. If you need this changed, you can do it in `main.tex`.

1.9 In Closing

You have reached the end of this mini-guide. You can now rename or overwrite this pdf file and begin writing your own `Chapter1.tex` and the rest of your thesis. The

easy work of setting up the structure and framework has been taken care of for you.
It's now your job to fill it out!

Good luck and have lots of fun!

Guide written by —
Sunil Patel: www.sunilpatel.co.uk
Vel: LaTeXTemplates.com

Chapter 2

Introduction to Machine Learning

2.1 Machine Learning Basics

Deep learning is a specific kind of machine learning. To understand deep learning well, one must have a solid understanding of machine learning. Machine learning is a form of applied statistics with an increased emphasis on the use of computers to statistically estimate complicated functions and a decreased emphasis on proving confidence intervals around these functions.

(Mitchell, 1997) describes a machine learning a task as: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E ."

2.1.1 The task T

In this definition of "task", the process of learning itself is not the task. For example, if we want a program to learn to distinguish dogs and cats, then distinguishing dogs and cats is the task.

Machine learning tasks are usually described in terms of how the machine learning system should process an example. An example is a collection of features that have been quantitatively measured from some object or event that we want the machine learning system to process. We typically represent an example as a vector $x \in \mathbb{R}^n$ where each entry x_i of the vector is a feature. For example, the features of an image are usually the values of the pixels in the image. Some typical examples of tasks are:

- **Classification:** Classification tasks ask the computer program to specify in which category does one input vector belong. Formally, this is equivalent to approximate a function $f : \mathbb{R}^n \mapsto \{1, 2, \dots, k\}$.
- **Regression:** In this type of task, the computer program is asked to predict a numerical value given some input. This consists in approximating a function $f : \mathbb{R}^n \mapsto \mathbb{R}$.
- **Transcription:** In this type of task, the machine learning system is asked to observe an unstructured representation of some kind of data and transcribe the information into text. For example, in optical character recognition (OCR), the computer program is shown an image containing text and is asked to return this text in the form of a sequence of characters.
- **Denoising:** In this type of task, the machine learning algorithm is given an input a corrupted example $\tilde{x} \in \mathbb{R}^n$ obtained by an unknown corruption process from a clean example $x \in \mathbb{R}^n$.

2.1.2 The Performance Measure, P

To evaluate how good our machine learning algorithm is, we must design quantitative measures of its performance. For tasks such as classification, we often measure the **accuracy** of the model, which is just the proportion of examples for which the algorithm gives the correct output. (similarly, the **error rate** is the proportion of examples for which the algorithm gives an incorrect output).

Usually, we are interested in seeing how well an algorithm performs on data it has not seen before, since this will determine how well it will work "in the real world". To determine this, we usually evaluate P using a **test set** of our data, which we have previously split from the **train set**.

2.1.3 The Experience, E

Machine learning algorithms can be understood as being allowed to experience an entire **dataset**. A dataset is a collection of many **examples**.

Generally, machine learning algorithms can be categorized as **supervised** or **unsupervised**.

- **Unsupervised**: The algorithm experiences a dataset containing features and has to learn some structure or pattern from it.
- **Supervised**: The algorithm experiences a dataset containing features, but each example is also associated with a **label** or **target**. We will generally refer the examples as \mathbf{x} and to the targets as \mathbf{y} . When we talk about a specific example and target, we will use superindices: $\mathbf{x}^{(i)}$ and $\mathbf{y}^{(i)}$.

Throughout this work, we will focus mainly on supervised learning, i. e. each example has $\mathbf{x}^{(i)}$ a corresponding target $\mathbf{y}^{(i)}$.

Some machine learning algorithms do not just experience a dataset. For example, **reinforcement learning** algorithms interact with an environment, so there is a feedback loop between the algorithm and its experiences. An example of reinforcement learning algorithm is AlphaGo Zero from Google (Silver et al., 2017), an algorithm that learned to play the game of Go from scratch better than any human.

2.2 Capacity, Overfitting and Underfitting

The central challenge of machine learning is that our algorithm must perform well on unseen examples. This is called **generalization**.

Typically, when training a machine learning model, we compute some error measure over our training set, called **training error**. This is simply an optimization problem. In machine learning, we want the **test error**, computed over a test set which the algorithm has never seen to be low as well. The test error will always be greater or equal than the training error.

The factors to determine how well a machine learning algorithm will perform are its ability to make the training error small, and to make the gap between the training error and the test (**generalization gap**) error small. These two factors correspond to **underfitting** (when the training error is too high) and **overfitting** (when the generalization gap is too high).

Whether a model is going to underfit or overfit depends on its **capacity**, which correspond to the capability of a model to approximate a wide variety of functions. Models with low capacity cannot fit the training set. Models with high capacity will overfit the training set, and the generalization gap will be large.

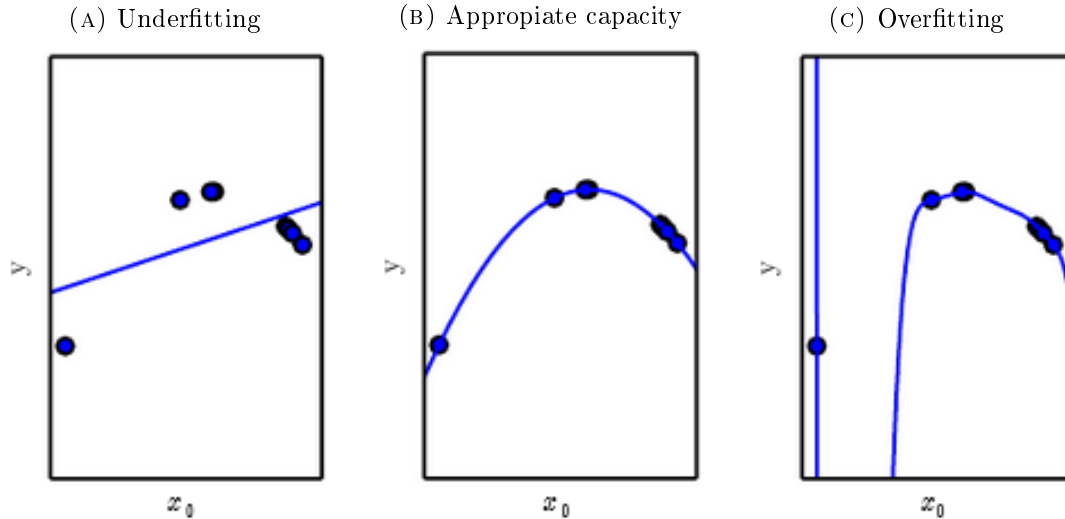


FIGURE 2.1: Here we have three examples of a regression. The training data was generated synthetically, by randomly sampling x values and choosing y deterministically by evaluating a quadratic function. (A) A linear function cannot capture the curvature and underfits the data. (B) A quadratic function fits to the data and generalizes well to unseen points. It does not suffer from overfitting or underfitting. (C) A polynomial of degree 9 overfits the data and does not generalize well to unseen points (Goodfellow, Bengio, and Courville, 2016).

2.2.1 Regularization

Regularization is a technique to avoid overfitting, and thus reduce the generalization gap. For example, in figure 2.1, we could "penalize" our algorithm for having large parameters. This would reduce the algorithm capacity (it would increase the training error), but it would help avoid overfitting (decreasing the generalization gap).

There are many ways to **regularize**, and which is the correct way depends on the task we want our algorithm to solve.

2.2.2 The No Free Lunch Theorem

The no free lunch theorem (Wolpert, 1996) says that, averaged over all possible data, every classification algorithm has the same error rate when classifying previously unobserved points. This means that no machine learning algorithm is better than any other in all possible tasks.

Fortunately, this results only holds when we average over all data-generating distributions. This means that if we make assumptions about the distributions that we will find in the real world, we can design an algorithm that does well in these distributions.

The no free lunch theorem implies that we must design our algorithm to work well only on a single task. This is done by assuming building hypothesis into the algorithm. For example, in a linear regression will work well if the outputs come from a linear function applied to the inputs, but will not work if the thata is of the form $y = \sin(x)$.

2.3 Gradient-Based Optimization

Many machine learning algorithms and nearly all deep learning algorithms use **Stochastic Gradient Descent (SGD)**, which is an extension from the classical Gradient Descent.

2.3.1 Gradient Descent

Generally, for an algorithm to learn a task it has to optimize some **objective function** or **loss function**. In general, this function will be a measure of the error, and we will want to minimize it with respect to the parameters of our algorithm.

It is well known that the gradient of a function gives the direction of maximum ascent, so, for small enough ϵ , $f(\mathbf{x} - \epsilon \nabla f(\mathbf{x}))$ will be smaller than $f(\mathbf{x})$. We can iterate this process, which will reduce $f(\mathbf{x})$ at each step. This technique is called gradient descent (Cauchy, 1847).

In machine learning, the ϵ parameter is called learning rate, as it symbolizes the speed at which the algorithm learns. Choosing the right learning rate is crucial, since a learning rate too low will make the algorithm learn too slowly, and a learning rate too high will make the gradient descent diverge and not find a minimum.

If L is the loss function that we want to minimize with respect to some parameters θ , and \mathbf{x} and \mathbf{y} are the examples with their corresponding labels, and \mathbf{f} is the function that we are approximating, the update rule will be:

$$\theta \leftarrow \theta - \epsilon \nabla_{\theta} L(\mathbf{f}(\mathbf{x}; \theta), \mathbf{y}) \quad (2.1)$$

2.3.2 Stochastic Gradient Descent

A problem in machine learning is that large datasets give better results, but they are also computationally expensive.

Usually, the loss function in a machine learning algorithm decomposes in the sum of some per-example loss function:

$$L_{\text{total}}(\mathbf{f}(\mathbf{x}; \theta), \mathbf{y}) = \frac{1}{m} \sum_{i=1}^m L(\mathbf{f}(\mathbf{x}^{(i)}; \theta), \mathbf{y}^{(i)}) \quad (2.2)$$

If the training examples grow to billions, the cost to make a single gradient step becomes too large.

The idea behind SGD is that the gradient is just an expected value of the largest descent direction, and it can be estimated with a small set of examples. On each step we can sample a **minibatch** of examples $\mathbb{B} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(m')}\}$ from the whole dataset, and estimate $\nabla f(x)$ using only these examples. The idea is that m' is fixed even though the size of the dataset can grow.

$$\nabla_{\theta} L_{\text{total}}(\mathbf{f}(\mathbf{x}; \theta), \mathbf{y}) \approx \nabla_{\theta} \frac{1}{m'} \sum_{i=1}^{m'} L(\mathbf{f}(\mathbf{x}^{(i)}; \theta), \mathbf{y}^{(i)}) \quad (2.3)$$

2.3.3 Momentum

Even though SGD works well, it can sometimes be slow. The method of momentum is designed to accelerate learning, especially in the face of high curvature, small but consistent gradients or noisy gradients.

The momentum algorithm introduces a variable v that plays the role of velocity. The name momentum derives from the physical analogy. The new update rule is given by:

$$\mathbf{v} \leftarrow \alpha \mathbf{v} - \epsilon \nabla_{\boldsymbol{\theta}} \frac{1}{m'} \sum_{i=1}^{m'} L(\mathbf{f}(\mathbf{x}^{(i)}; \boldsymbol{\theta}), \mathbf{y}^{(i)}) \quad (2.4)$$

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \mathbf{v} \quad (2.5)$$

Here, α determines how quickly the contribution from previous gradients decays.

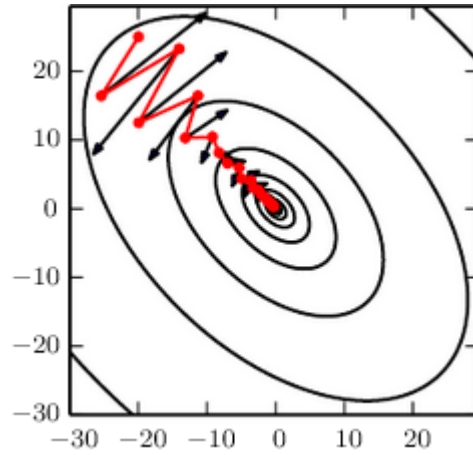


FIGURE 2.2: Here we can see that the steps (in red) that the momentum takes go in the direction of the minimum, while the gradient (in black) points in other directions (Goodfellow, Bengio, and Courville, 2016).

Chapter 3

Introduction to Deep Learning

3.1 Main Section 1

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

3.1.1 Subsection 1

Nunc posuere quam at lectus tristique eu ultrices augue venenatis. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Aliquam erat volutpat. Vivamus sodales tortor eget quam adipiscing in vulputate ante ullamcorper. Sed eros ante, lacinia et sollicitudin et, aliquam sit amet augue. In hac habitasse platea dictumst.

3.1.2 Subsection 2

Morbi rutrum odio eget arcu adipiscing sodales. Aenean et purus a est pulvinar pellentesque. Cras in elit neque, quis varius elit. Phasellus fringilla, nibh eu tempus venenatis, dolor elit posuere quam, quis adipiscing urna leo nec orci. Sed nec nulla auctor odio aliquet consequat. Ut nec nulla in ante ullamcorper aliquam at sed dolor. Phasellus fermentum magna in augue gravida cursus. Cras sed pretium lorem. Pellentesque eget ornare odio. Proin accumsan, massa viverra cursus pharetra, ipsum nisi lobortis velit, a malesuada dolor lorem eu neque.

3.2 Main Section 2

Sed ullamcorper quam eu nisl interdum at interdum enim egestas. Aliquam placerat justo sed lectus lobortis ut porta nisl porttitor. Vestibulum mi dolor, lacinia molestie gravida at, tempus vitae ligula. Donec eget quam sapien, in viverra eros. Donec pellentesque justo a massa fringilla non vestibulum metus vestibulum. Vestibulum in orci quis felis tempor lacinia. Vivamus ornare ultrices facilisis. Ut hendrerit volutpat vulputate. Morbi condimentum venenatis augue, id porta ipsum vulputate in. Curabitur luctus tempus justo. Vestibulum risus lectus, adipiscing nec condimentum quis, condimentum nec nisl. Aliquam dictum sagittis velit sed iaculis. Morbi tristique augue sit amet nulla pulvinar id facilisis ligula mollis. Nam elit libero, tincidunt ut aliquam at, molestie in quam. Aenean rhoncus vehicula hendrerit.

Chapter 4

Adaptive Computation Time

4.1 Main Section 1

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

4.1.1 Subsection 1

Nunc posuere quam at lectus tristique eu ultrices augue venenatis. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Aliquam erat volutpat. Vivamus sodales tortor eget quam adipiscing in vulputate ante ullamcorper. Sed eros ante, lacinia et sollicitudin et, aliquam sit amet augue. In hac habitasse platea dictumst.

4.1.2 Subsection 2

Morbi rutrum odio eget arcu adipiscing sodales. Aenean et purus a est pulvinar pellentesque. Cras in elit neque, quis varius elit. Phasellus fringilla, nibh eu tempus venenatis, dolor elit posuere quam, quis adipiscing urna leo nec orci. Sed nec nulla auctor odio aliquet consequat. Ut nec nulla in ante ullamcorper aliquam at sed dolor. Phasellus fermentum magna in augue gravida cursus. Cras sed pretium lorem. Pellentesque eget ornare odio. Proin accumsan, massa viverra cursus pharetra, ipsum nisi lobortis velit, a malesuada dolor lorem eu neque.

4.2 Main Section 2

Sed ullamcorper quam eu nisl interdum at interdum enim egestas. Aliquam placerat justo sed lectus lobortis ut porta nisl porttitor. Vestibulum mi dolor, lacinia molestie gravida at, tempus vitae ligula. Donec eget quam sapien, in viverra eros. Donec pellentesque justo a massa fringilla non vestibulum metus vestibulum. Vestibulum in orci quis felis tempor lacinia. Vivamus ornare ultrices facilisis. Ut hendrerit volutpat vulputate. Morbi condimentum venenatis augue, id porta ipsum vulputate in. Curabitur luctus tempus justo. Vestibulum risus lectus, adipiscing nec condimentum quis, condimentum nec nisl. Aliquam dictum sagittis velit sed iaculis. Morbi tristique augue sit amet nulla pulvinar id facilisis ligula mollis. Nam elit libero, tincidunt ut aliquam at, molestie in quam. Aenean rhoncus vehicula hendrerit.

Chapter 5

Experiments

All experiments are run with ADAM optimizer and learning rate of 10^{-4}

5.1 Addition

The addition task presents the network with a input sequence of 1 to 5 size 50 input vectors. Each vector represents a D digit number, where D is drawn randomly from 1 to 5, and each digit is drawn randomly from 0 to 9. The first $10D$ elements of the vector are a concatenation of one-hot encodings of the D digits in the number, and the remainder of the vector is set to 0. The required output is the cumulative sum of all inputs up to the current one, represented as a set of 6 simultaneous classifications for the 6 possible digits in the sum. There is no target for the first vector in the sequence, as no sums have yet been calculated. Because the previous sum must be carried over by the network, this task again requires the internal state of the network to remain coherent. Each classification is modelled by a size 11 softmax, where the first 10 classes are the digits and the 11th is a special marker used to indicate that the number is complete.

The network was single-layer LSTM with 512 memory cells. The loss function was the joint cross-entropy of all 6 targets at each time-step where targets were present and the minibatch size was 32. The maximum ponder M was set to 20 for this task, as it was found that some networks had very high ponder times early in training

5.1.1 Subsection 1

Nunc posuere quam at lectus tristique eu ultrices augue venenatis. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Aliquam erat volutpat. Vivamus sodales tortor eget quam adipiscing in vulputate ante ullamcorper. Sed eros ante, lacinia et sollicitudin et, aliquam sit amet augue. In hac habitasse platea dictumst.

5.1.2 Subsection 2

Morbi rutrum odio eget arcu adipiscing sodales. Aenean et purus a est pulvinar pellentesque. Cras in elit neque, quis varius elit. Phasellus fringilla, nibh eu tempus venenatis, dolor elit posuere quam, quis adipiscing urna leo nec orci. Sed nec nulla auctor odio aliquet consequat. Ut nec nulla in ante ullamcorper aliquam at sed dolor. Phasellus fermentum magna in augue gravida cursus. Cras sed pretium lorem. Pellentesque eget ornare odio. Proin accumsan, massa viverra cursus pharetra, ipsum nisi lobortis velit, a malesuada dolor lorem eu neque.

5.2 Parity

Sed ullamcorper quam eu nisl interdum at interdum enim egestas. Aliquam placerat justo sed lectus lobortis ut porta nisl porttitor. Vestibulum mi dolor, lacinia molestie gravida at, tempus vitae ligula. Donec eget quam sapien, in viverra eros. Donec pellentesque justo a massa fringilla non vestibulum metus vestibulum. Vestibulum in orci quis felis tempor lacinia. Vivamus ornare ultrices facilisis. Ut hendrerit volutpat vulputate. Morbi condimentum venenatis augue, id porta ipsum vulputate in. Curabitur luctus tempus justo. Vestibulum risus lectus, adipiscing nec condimentum quis, condimentum nec nisl. Aliquam dictum sagittis velit sed iaculis. Morbi tristique augue sit amet nulla pulvinar id facilisis ligula mollis. Nam elit libero, tincidunt ut aliquam at, molestie in quam. Aenean rhoncus vehicula hendrerit.

Chapter 6

Conclusions

6.1 Main Section 1

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

6.1.1 Subsection 1

Nunc posuere quam at lectus tristique eu ultrices augue venenatis. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Aliquam erat volutpat. Vivamus sodales tortor eget quam adipiscing in vulputate ante ullamcorper. Sed eros ante, lacinia et sollicitudin et, aliquam sit amet augue. In hac habitasse platea dictumst.

6.1.2 Subsection 2

Morbi rutrum odio eget arcu adipiscing sodales. Aenean et purus a est pulvinar pellentesque. Cras in elit neque, quis varius elit. Phasellus fringilla, nibh eu tempus venenatis, dolor elit posuere quam, quis adipiscing urna leo nec orci. Sed nec nulla auctor odio aliquet consequat. Ut nec nulla in ante ullamcorper aliquam at sed dolor. Phasellus fermentum magna in augue gravida cursus. Cras sed pretium lorem. Pellentesque eget ornare odio. Proin accumsan, massa viverra cursus pharetra, ipsum nisi lobortis velit, a malesuada dolor lorem eu neque.

6.2 Main Section 2

Sed ullamcorper quam eu nisl interdum at interdum enim egestas. Aliquam placerat justo sed lectus lobortis ut porta nisl porttitor. Vestibulum mi dolor, lacinia molestie gravida at, tempus vitae ligula. Donec eget quam sapien, in viverra eros. Donec pellentesque justo a massa fringilla non vestibulum metus vestibulum. Vestibulum in orci quis felis tempor lacinia. Vivamus ornare ultrices facilisis. Ut hendrerit volutpat vulputate. Morbi condimentum venenatis augue, id porta ipsum vulputate in. Curabitur luctus tempus justo. Vestibulum risus lectus, adipiscing nec condimentum quis, condimentum nec nisl. Aliquam dictum sagittis velit sed iaculis. Morbi tristique augue sit amet nulla pulvinar id facilisis ligula mollis. Nam elit libero, tincidunt ut aliquam at, molestie in quam. Aenean rhoncus vehicula hendrerit.

Appendix A

Frequently Asked Questions

A.1 How do I change the colors of links?

The color of links can be changed to your liking using:

```
\hypersetup{urlcolor=red}, or  
\hypersetup{citecolor=green}, or  
\hypersetup{allcolor=blue}.
```

If you want to completely hide the links, you can use:

```
\hypersetup{allcolors=}, or even better:  
\hypersetup{hidelinks}.
```

If you want to have obvious links in the PDF but not the printed text, use:

```
\hypersetup{colorlinks=false}.
```


Bibliography

- Campos, V. et al. (2017). “Skip RNN: Learning to Skip State Updates in Recurrent Neural Networks”. In: *ArXiv e-prints*. arXiv: [1708.06834 \[cs.AI\]](#).
- Cauchy, Augustin (1847). “Méthode générale pour la résolution des systemes d’équations simultanées”. In:
- Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press.
- Graves, A. (2016). “Adaptive Computation Time for Recurrent Neural Networks”. In: *ArXiv e-prints*. arXiv: [1603.08983](#).
- Mitchell, Tom (1997). *Machine Learning*.
- Silver, David et al. (2017). “Mastering the game of Go without human knowledge”. In: *Nature* 550.7676, pp. 354–359. ISSN: 0028-0836. DOI: [10.1038/nature24270](#). URL: <http://doi.org/10.1038/nature24270>.
- Wolpert, David H. (1996). “The Lack of A Priori Distinctions Between Learning Algorithms”. In: *Neural Computation* 8.7, pp. 1341–1390. DOI: [10.1162/neco.1996.8.7.1341](#). eprint: <https://doi.org/10.1162/neco.1996.8.7.1341>. URL: <https://doi.org/10.1162/neco.1996.8.7.1341>.