



**Department of Mathematics and Geosciences**  
**Master Degree in Data Science and Scientific**  
**Computing**

The Dynamics of Learning Beyond Stochastic Gradient  
Descent

**Supervisor:**  
Prof. Sebastian Goldt

**Candidate:**  
Doğan Can Demirbilek

---

**Academic Year 2020/2021**

*To my family*

# Acknowledgements

This is where you will normally thank your advisor, colleagues, family and friends, as well as funding and institutional support. In our case, we will give our praises to the people who developed the ideas and tools that allow us to push open science a little step forward by writing plain-text, transparent, and reproducible theses in R Markdown.

We must be grateful to John Gruber for inventing the original version of Markdown, to John MacFarlane for creating Pandoc (<http://pandoc.org>) which converts Markdown to a large number of output formats, and to Yihui Xie for creating `knitr` which introduced R Markdown as a way of embedding code in Markdown documents, and `bookdown` which added tools for technical and longer-form writing.

Special thanks to [Chester Ismay](#), who created the `thesisdown` package that helped many a PhD student write their theses in R Markdown. And a very special thanks to John McManigle, whose adaption of Sam Evans' adaptation of Keith Gillow's original maths template for writing an Oxford University DPhil thesis in LaTeX provided the template that I in turn adapted for R Markdown.

Finally, profuse thanks to JJ Allaire, the founder and CEO of [RStudio](#), and Hadley Wickham, the mastermind of the tidyverse without whom we'd all just given up and done data science in Python instead. Thanks for making data science easier, more accessible, and more fun for us all.

Ulrik Lyngs  
Linacre College, Oxford  
2 December 2018

# Abstract

This *R Markdown* template is for writing an Oxford University thesis. The template is built using Yihui Xie's `bookdown` package, with heavy inspiration from Chester Ismay's `thesisdown` and the `OxThesis` L<sup>A</sup>T<sub>E</sub>X template (most recently adapted by John McManigle).

This template's sample content include illustrations of how to write a thesis in R Markdown, and largely follows the structure from [this R Markdown workshop](#).

Congratulations for taking a step further into the lands of open, reproducible science by writing your thesis using a tool that allows you to transparently include tables and dynamically generated plots directly from the underlying data. Hip hooray!

# Contents

<b>List of Figures</b>	<b>vi</b>
<b>List of Tables</b>	<b>vii</b>
<b>List of Abbreviations</b>	<b>viii</b>
<b>Introduction</b>	<b>1</b>
<b>1 Theoretical Foundations</b>	<b>4</b>
1.1 Backpropagation . . . . .	4
1.2 Direct Feedback Alignment . . . . .	8
1.3 Lazy Methods . . . . .	11
1.4 Optimizers . . . . .	13
<b>2 Learning Problems</b>	<b>19</b>
2.1 Parity Learning Problem . . . . .	19
2.2 Random Data . . . . .	21
<b>3 Citations, cross-references, and collaboration</b>	<b>22</b>
3.1 Citations . . . . .	22
3.2 Cross-referencing . . . . .	25
3.3 Collaborative writing . . . . .	28
3.4 Additional resources . . . . .	29
<b>4 Tables</b>	<b>30</b>
4.1 Making LaTeX tables play nice . . . . .	30
<b>5 Customisations and extensions</b>	<b>44</b>
5.1 Front matter . . . . .	45
5.2 Shorten running header (PDF) . . . . .	46
5.3 Unnumbered chapters . . . . .	46

## Contents

5.4	Beginning chapters with quotes (PDF) . . . . .	46
5.5	Highlighting corrections (HTML & PDF) . . . . .	47
5.6	Apply custom font color and highlighting to text (HTML & PDF) .	48
5.7	Including another paper in your thesis - embed a PDF document . .	49
5.8	Including another paper in your thesis - R Markdown child document	53
5.9	Customizing referencing . . . . .	57
5.10	Customizing the page headers and footers (PDF) . . . . .	59
5.11	Diving in to the OxThesis LaTeX template (PDF) . . . . .	60
5.12	Customising to a different university . . . . .	60
<b>6</b>	<b>Troubleshooting</b>	<b>62</b>
6.1	Error: Failed to build the bibliography via biber . . . . .	62
	<b>Conclusion</b>	<b>64</b>
	More info . . . . .	64
	<b>Appendices</b>	
<b>A</b>	<b>Backpropagation with Binary Cross-Entropy</b>	<b>66</b>
<b>B</b>	<b>Reproducibility</b>	<b>69</b>
	<b>References</b>	<b>70</b>

# List of Figures

1.1	Error Transportation in Backpropagation . . . . .	6
1.2	Error Transportation in Direct Feedback Alignment . . . . .	9
2.1	Reproduced MNIST parity experiment result from [3] . . . . .	21
3.1	The ‘citr’ add-in . . . . .	25
3.2	A marvel-lous meme . . . . .	27
4.1	Font sizes in LaTeX . . . . .	41

# List of Tables

3.1	Stopping cars . . . . .	27
-----	-------------------------	----



# List of Abbreviations

<b>ANN</b>	. . . . .	Artificial Neural Network.
<b>BP</b>	. . . . .	Backpropagation.
<b>DFA</b>	. . . . .	Direct Feedback Alignment.
<b>BCE</b>	. . . . .	Binary Cross Entropy.
<b>FA</b>	. . . . .	Feedback Alignment.
<b>NTK</b>	. . . . .	Neural Tangent Kernel.
<b>t-SNE</b>	. . . . .	t-Distributed Stochastic Neighbor Embedding.
<b>SGD</b>	. . . . .	Stochastic Gradient Descent.
<b>NAG</b>	. . . . .	Nesterov Accelerated Gradient.

# Introduction

Artificial neural networks (ANNs) are a collection of connected computational nodes inspired by biological neural networks. Each connection can transmit a helpful signal to another computational node like synapses in a brain. ANNs demonstrated colossal advancements in the last decades. Thanks to these advancements, it is possible to solve complex problems in computer vision, speech recognition, and natural language processing within a reasonable amount of time and with satisfactory performance. These advancements were actualized through an old but robust algorithm called backpropagation (BP). BP is a training algorithm for ANNs based on repeatedly adjusting network weights to minimize the difference (loss) between the output of the network and the ground truth [1].

Although nowadays BP is the workhorse algorithm for training ANNs, it has some drawbacks, and it is not the only alternative. Recent studies offered different algorithms to train ANNs by addressing these drawbacks. These algorithms have other properties and principles than BP. Some of them are competitive with BP, or they even outperform the BP in terms of performance or convergence speed for specific problems.

This thesis investigates the learning structures through BP and one of the alternative algorithm called direct feedback alignment (DFA) on the particular problem. Unlike BP, the error is propagated through a fixed random matrix instead of the layers' weights in DFA. Then network learns how to make this feedback useful [2]. Due to this error propagation mechanism, DFA is considered more biologically plausible than BP, and it opens the gate of parallelism in the training phase of ANNs.

The problem at hand is known as the parity learning problem. Previous results showed that these parities are learnable by BP and lazy methods in a more simple

## *Introduction*

setting, whereas it is only learnable by BP in a more complex setting [3]. That is why it is intriguing to test alternative algorithms on this problem to understand their learning dynamics and capabilities.

The experiment results might lead us to three possible outcomes. First, we might acquire a similar performance as BP. If it is the case, it would be beneficial to test DFA and BP on a more challenging problem for further studies. Second, there might be a gap between BP and DFA then it would be intriguing to understand where the difference is coming from and how we can close this gap. Third, the alternative algorithm might not even learn, and in this case, it is interesting to ask what makes a problem learnable by BP but not DFA. In all cases, results should help to understand the dynamics of learning of both methods.

For applying BP and DFA in a more realistic setting, experiments are performed on the MNIST dataset by imitating the parity learning problem. After putting DFA to this frame, the reason behind the results is interpreted, and possible improvements are motivated and implemented.

Chapter 1 constructs the theoretical bases of the algorithms that are used for the experiments. These bases are composed of simple definitions, mathematical foundations, and the drawbacks of the algorithms. They help to dig deeper into the learning structures of the training algorithms. It is expected to have more control over their learning behaviors by tweaking components of these foundations. Also, it is beneficial to have these theoretical bases for acquiring a better understanding of the further interventions. Moreover, these theoretical foundations are used to implement the algorithms from scratch to use in experiments.

Chapter 2 introduces the parity learning problem at hand. First, the formal definition of the problem is demonstrated then how the problem is imitated by using the MNIST dataset is explained in detail. This part is also highly correlated with the training phase of the algorithms. Later experiment results from [3] are replicated to have a concrete picture of previous studies.

Chapter 3 presents results of the experiments. After having the same results from the previous studies for the parity problem, DFA is tested on the same problem.

## *Introduction*

Later, a couple of experiments are performed in DFA, such as trying different random matrices and optimizers for the same problem. Moreover, hidden representations of BP and DFA are observed to test if networks memorize the data or learn the digits, particularly in the parity learning problem.

Chapter 4 wraps up the findings from experiments and creates a path for future studies.

# 1

## Theoretical Foundations

### Contents

---

<b>1.1</b>	<b>Backpropagation</b>	<b>4</b>
1.1.1	Drawbacks of BP	6
<b>1.2</b>	<b>Direct Feedback Alignment</b>	<b>8</b>
<b>1.3</b>	<b>Lazy Methods</b>	<b>11</b>
1.3.1	Neural Tangent Kernel	12
1.3.2	Random features	12
<b>1.4</b>	<b>Optimizers</b>	<b>13</b>
1.4.1	Gradient Descent	13
1.4.2	Adaptive Methods	15

---

### 1.1 Backpropagation

BP is one of the first algorithms that show ANNs could learn hidden representations well. Numerous studies showed that ANNs trained with BP could capture similar information as biological neural networks (e.g., specific nodes learn the edges, corners). We need three components for BP, a dataset composed of input-output pairs, a network consisting of parameters (weights and biases), and it allows the input to flow through the network to have output. We need a loss function to measure the difference between the output of the network and the ground truth that we have from the dataset.

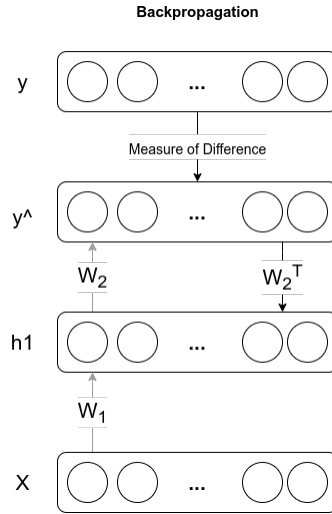
## 1. Theoretical Foundations

The main goal of BP is computing the gradients of the loss function (a measure of difference) concerning the parameters of neural networks by using the chain rule. These gradients show how much the parameter needs to change (positively or negatively) to minimize the loss function. After efficiently calculating the gradients, we can nudge the network parameters using gradient descent or its variants.

Although BP is an older idea, it earned popularity with [1] because it presented how BP can make a network to learn the representations. After this popularity, the community published numerous practical and theoretical papers that investigated the dynamics of BP. It would be repeat and infeasible to show all the aspects again. However, for completeness and a smoother transition from BP to DFA, it is beneficial to have visual and mathematical explanations showing how the algorithms propagate the errors and the weights. For the mathematical foundations, a binary classification task will be demonstrated with binary cross-entropy loss as an example in appendix A. This example is not chosen arbitrarily. Indeed the parity problem that MNIST imitates is a binary classification problem. In addition to this, equations from appendix A are used to implement BP from scratch to have more control over the process. Then the exact implementation is modified to obtain DFA. The same set of steps are valid for different loss functions and activation functions. Only the calculations will be slightly different. The general idea is the same: obtaining the gradients by calculating the derivative of the loss function concerning the parameters.

In figure 1.1 we have a simple network with only a hidden layer that shows the error transportation configuration in BP.  $W_i$  are the weights,  $h_i$  are the output of the hidden layers that is denoted as  $i$ ,  $\hat{y}$  is the output of the network, and  $y$  is the ground truth, for the sake of simplicity, biases are not showed in this figure. It is important to note that in BP, the transpose of weight is propagated to calculate the gradients. In literature, this issue is known as the weight transport problem, and it is one of the most criticized disadvantages of BP.

## 1. Theoretical Foundations



**Figure 1.1:** Error Transportation in Backpropagation

### 1.1.1 Drawbacks of BP

We know that biological neurons inspired ANNs. However, recent studies showed that BP is not precisely how biological neurons learn [4]. This brings a term called biological plausibility of an algorithm that indicates the algorithm's consistency with existing biological, medical, and neuroscientific knowledge. That is why the community proposed many alternative algorithms by addressing these limitations of BP. In the light of this term, we can put in order the drawbacks of BP as the following:

- **Biological implausibility:**
  - The BP computation is purely linear whereas biological neurons interleave linear and non-linear operations.
  - BP needs precise knowledge of derivatives of the non-linearities at the operating point used in the corresponding feedforward computation on the feedforward path.
  - BP has to use exact symmetric weights of the feedforward connections.
  - Real neurons communicate by binary values (spikes), not by clean continuous values.

## 1. Theoretical Foundations

- The computation has to be precisely clocked to alternate between feedforward and BP phases.
- It is not clear where the output targets would come from. [4, 5]
- **Vanishing or Exploding Gradients**
- **Lack of Parallel Processing** [6]

Simple interventions may handle some of these drawbacks. For instance, implementing gradient clipping or different activation functions might solve exploding gradients and vanishing gradients. However, they may frequently happen in deeper networks, and they must be considered while training ANNs. On the other hand, some of the drawbacks can not be handled with superficial modifications. For instance, BP is a sequential process, and there are locking mechanisms (forward, backward and update) that ensure none of the processes is executed before its preceding completion. This makes BP infeasible for parallel processing because each execution has to wait for its preceding process. Hence deeper and larger networks' training can be computationally expensive.

Biological plausibility is significant because of a couple of reasons. We know that biological neurons inspired ANNs, and biological plausibility refers to consistency between BP and biological knowledge about the neurons of a brain. Hence, it is interesting to examine the dissimilarity or similarity among them. Besides, there is a field that is the intersection of neuroscience and deep learning, so it is natural to investigate the biological plausibility feature of the algorithms, especially for this field. Furthermore, even though nowadays ANNs might outperform the human brain in a specific task, we are still distant from fully mimicking it. In other words, most of the time, ANNs are very good on a task that they are trained in, but they are not diverse, and some kinds of attacks like adversarial ones can easily trick them. Investigating the learning dynamics of these algorithms may open the doors of diverse ANNs that are not specialized in a single task or make them more robust to attacks.



## 1. Theoretical Foundations

Alternative algorithms address some of the drawbacks of BP, and they propose a solution to them, but they also demonstrate a couple of them. However, these algorithms can be considered one or more steps closer to more biologically plausible and robust algorithms.

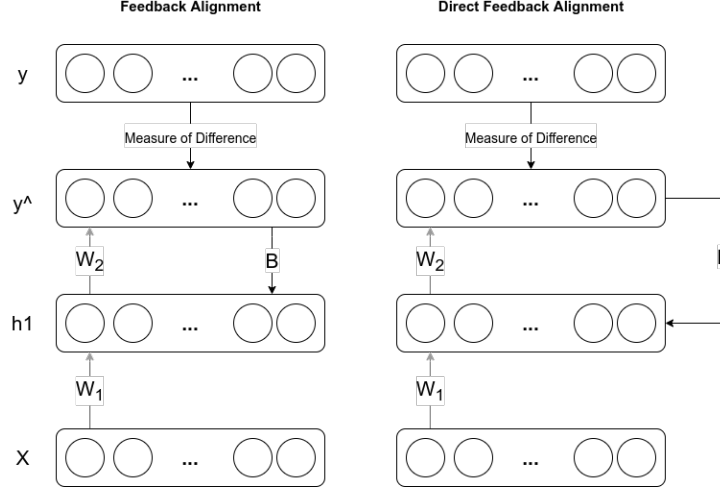
### 1.2 Direct Feedback Alignment

So far, we have seen how the error is propagated in BP sequentially through a network with the backward pass. Unlike BP, DFA uses a different way to propagate the error. This way uses a random matrix instead of the transpose of the weight matrix, which brings a solution to the weight transport problem. Before explaining how DFA works, it is better to investigate the feedback alignment (FA) algorithm since DFA is the extension of FA.

In [7], authors proved that precise symmetric weights are not required to obtain learning in ANNs. Without these matrices, BP-like learning can be obtained. Any random matrix under some conditions can provide the learning. Implicit dynamics in the standard forward weight updates encourage an alignment between weights and the random matrix. In other words, a random matrix pushes the network in roughly the same direction as BP would. They supported this hypothesis with some experiments on a linear problem and MNIST classification task. The empirical results demonstrate that FA successfully trains the network and has similar performance results as BP on these tasks.

Even though learning still occurs with random matrix and FA offers the solution to the weight transport problem, it does not provide any computational advantage. To extend DFA, we need to change the error propagation mechanism of FA slightly. In FA, the error is propagated through a random matrix, but the backward process is still sequential. DFA extends this idea and propagates the random matrix in parallel to each layer. In other words, DFA takes the loss and distribute it globally to all layers without requiring sequential step. It also creates an opportunity to parallelize the computation that might speed up the training process.

## 1. Theoretical Foundations



**Figure 1.2:** Error Transportation in Direct Feedback Alignment

In figure 1.2 we can see the error transportation configurations for FA and DFA. This figure is the same as the one in [2] but shows only one hidden layer. In fact, with only one hidden layer, FA and DFA are identical.

It is crucial to point out, BP and DFA have different learning dynamics. BP calculates the gradients that point to the steepest descent in the loss function space. On the other hand, FA and DFA provide a different update direction but still descending. It is still descending because empirical and theoretical results proved that the networks' weights align with the random matrix that leads to gradients alignment. Therefore the more alignment we have, the same direction would FA and DFA point as BP. Even though they have different update directions, since they are both descending, the results from [2, 7] showed that FA and DFA are as good as BP in terms of performance for specified tasks in these papers. In addition to this, ANNs trained with DFA show decent separation for labels as in BP's hidden representations of the layers. We can observe this from the t-distributed stochastic neighbor embedding (t-SNE) visualizations of the hidden layers' representations. t-SNE is a method visualizing high-dimensional data which tries to keep the neighbor property in lower dimensions.

Recently, a new study has been published which tests the applicability of DFA on modern deep learning tasks and architectures such as neural view synthesis, recommender systems, geometric learning, and natural language processing [8].

## 1. Theoretical Foundations

Because even though some of the alternative methods are competitive with BP in simple tasks like MNIST, they are not competitive or trainable on more complex tasks. Results showed that DFA successfully trains all these complex architectures with performance close to BP. This study supports that complex tasks can be solved without symmetric weight transport, proving that DFA is suitable for more challenging problems.

Let us use the same example as [A](#) to present how gradients are calculated in DFA. After having the mathematical foundations of BP, transition to DFA is relatively easy. The forward pass is the same as BP, whereas, in the backward pass, we need to replace the transpose of the weight matrix, which is used to calculate the gradients with the random matrix. Considering the same example, we have a simple binary classification task with binary cross-entropy loss, and our network has only one hidden layer. In this setting, gradients of the weights can be calculated as the following:

$$\frac{\partial BCE}{\partial w_2} = h_1^T (\hat{y} - y)$$

There is no change in the calculations of gradients of the last layer, whereas, for the hidden layer, we have:

$$\frac{\partial BCE}{\partial w_1} = (X)^T (\hat{y} - y) (B) \odot f'(a_1)$$

Please pay attention that  $w_2^T$  is replaced with the random matrix  $B$ . This means that we can obtain learning by changing either the random matrix or weight matrix. We know that in DFA,  $B$  is fixed, so the feedforward weights of the network will learn to make these signals useful by aligning with the BP's teaching signal.

Update rules are the same as BP, which means that gradient descent and its variants can be used.

$$\text{parameter} = \text{parameter} - \text{step size} \times \frac{\partial BCE}{\partial(\text{parameter})}$$

With this tiny modification, DFA brings a solution to some of the drawbacks of BP. Such as using exact symmetric weights of the feedforward connections (weight

### *1. Theoretical Foundations*

transport problem), lack of parallel processing (random matrix can be propagated in parallel), and it is less likely to suffer from vanishing or exploding gradients than BP. Eventually, it proposes a more biologically plausible training method. However, it is not the perfect solution either. Because it assumes there is a global feedback path to propagate the error that might be biologically implausible because feedback has to travel a long physical distance. It also suffers some of the drawbacks of BP. For instance, computation is still purely linear. We still need precise knowledge of derivatives of non-linearities. We still communicate by clean, continuous values, and it is unclear where the output targets would come from. Besides, DFA has an extra task to accomplish while training the ANN that aligns with BP’s weights, and a layer can not learn before its preceding layers are aligned. This might spawn performance concerns, and DFA might lag behind BP. Furthermore, DFA fails to train convolutional neural networks which dominate the computer vision tasks [9, 10]. Finally, unlike BP, DFA was not investigated on particular subjects like adversarial attacks and interpretability by the community. This leaves some question marks about the robustness of DFA.

## **1.3 Lazy Methods**

Theoretical results present that especially over-parameterized ANNs (not limited to these networks) trained with gradient-based methods can reach zero training loss with their parameters barely changing. The term lazy does not refer to the poor property of methods, whereas they are called lazy methods because their parameters hardly move during training [11].

Lazy methods are not at the center of the experiments. Hence, detailed explanations of these methods are out of scope in this study, but they have been presented in [3], and they fail to learn the parities in a more complex setting. Hence, they are implemented too for completeness, and it is essential to embody at least a simple definition of them and how they are practically implemented.

## 1. Theoretical Foundations

### 1.3.1 Neural Tangent Kernel

Studies showed that neural networks under some conditions are equivalent to gaussian process, and they mathematically approximate the kernel machines if they are trained with gradient descent [12, 13]. Authors of [14] proved that during the training phase, ANNs follow the kernel gradient of the functional loss concerning a new kernel. They named this kernel a Neural Tangent Kernel (NTK). In other words, NTK is a kernel that describes the evolution of an ANN during the learning phase, and it is beneficial to explain the training of ANNs in function space rather than parameters space. It allows us to work with infinite width neural networks using the kernel trick, and it helps us understand the dynamics of learning and interference.

Empirical results demonstrated that the NTK regime performs worse than BP on standard tasks like MNIST. However, NTK is still worth investigating further to understand ANNs' training dynamics since it brings a new perspective on the training phase.

Basic practical implementation of NTK is obtained with three steps. Initially, an extra layer is created with the exact dimensions of the first layer. The second in the forward pass concatenation of these two layers' parameters are given as input to the gated linear unit with 1. Lastly, in the parameters update phase, the extra layer is not considered. With these adjustments, we decoupled the gating from the linearity of the ReLU, and we kept the gates fixed during training.

### 1.3.2 Random features

Standard random features are where first layer weights are initialized randomly by following a distribution and the train only the second layer. These mechanisms are particularly good at approximating kernels. They are preferred because kernel machines might take too much time to train if the data size is big. In **gaussian features** case, we initialize the first layer weights using gaussian distribution. In contrast, in **ReLU features** and **linear features**, we initialize the first layer

## 1. Theoretical Foundations

weights uniformly but, for linear features, non-linear activation functions are not used in the forward pass.

### 1.4 Optimizers

Up to this point, we only mentioned how we could use gradient descent and its variants to update the weights of a network superficially. This part is worth further investigation because many variants provide better convergence properties to find the minimum of the loss function. We may take advantage of these methods to have better performance or faster convergence for BP and DFA. These methods may spawn a significant impact on convergence speed and overall performance. As a reference to the following methods, mostly [15] is used, also the structure of this part and mathematical notations are adapted from the same paper, it is an excellent overview for the optimizers, and it reviews their advantages as well as drawbacks.

#### 1.4.1 Gradient Descent

Gradient descent is a first-order iterative optimization algorithm. It is the most used algorithm to optimize neural networks. It has three variants that depend on how much data we use to compute the gradients. **Batch gradient descent** computes the gradients for the entire dataset and performs only one update. **Stochastic gradient descent** (SGD), in contrast, calculates gradients for each training example and performs parameter update for each of them. Lastly, **mini-batch gradient descent** calculates the gradients of mini-batches and performs updates for each mini-batches. Gradient descent is infeasible to implement for the datasets that do not fit in the memory. In contrast, SGD performs too frequent updates, spawning high variance in parameters that cause fluctuation in the loss function. SGD provides the same convergence properties as batch gradient descent if the learning rate periodically decreases through iterations. For our experiments, we used mini-batch gradient descent, which takes the best of two methods. Most of the implementations use SGD

## 1. Theoretical Foundations

term instead of mini-batch gradient descent. The same tradition will be followed in this study too. Update rule of mini-batch gradient descent is the following:

$$\theta_{t+1} = \theta_t - \eta \cdot \nabla_{\theta} J\left(\theta; x^{(i:i+n)}; y^{(i:i+n)}\right)$$

where  $\theta$  is the parameters of the network,  $\eta$  is the learning rate or step size,  $\nabla_{\theta}$  is the gradients of the parameters and  $J\left(\theta; x^{(i:i+n)}; y^{(i:i+n)}\right)$  is the loss function for mini-batch  $i$  to  $i + n$ .

There are a couple of challenges in SGD because it doesn't always guarantee good convergence:

- Choosing a proper learning rate is intricate. Small learning rates may take too long to converge, whereas big learning rates may spawn loss function fluctuations and even diverge.
- SGD does not guarantee the global minimum. It can easily be stuck in the local minimum for highly non-convex loss functions that are standard for deep learning tasks.
- Same learning rate is applied to all parameters, but we may want to update the parameter by their frequencies.
- Convergence is strongly dependent on where the initial step starts. Unfortunate initializations may never reach the global minimum.

### Momentum

SGD has difficulties finding the direction in valleys because the gradients on these areas will be either zero or very close to zero, so it will slow down and make hesitant progress. These areas are prevalent around the local minimum. Momentum is an idea that dampens the oscillations in the relevant direction. It is accomplished by adding a fraction  $\gamma$  of the update vector of the past time step. This fraction is usually set to 0.9. This term usually leads to faster convergence and speeds up the iterations.

$$v_t = \gamma v_{t-1} + \eta \nabla_{\theta} J(\theta)$$

$$\theta_{t+1} = \theta_t - v_t$$

## 1. Theoretical Foundations

However, momentum follows the direction of the gradients blindly, **nesterov accelerated gradient** (NAG) is a way of giving our method to intuition by approximating the next position of the parameters with  $\theta - \gamma v_{t-1}$ , with this we hope to slow down before the hill slopes up. In other words, first, as in the momentum method, we make a big jump in the direction of previous gradients, then we measure the gradients where we end up and make a correction. The new update rule becomes:

$$\begin{aligned}v_t &= \gamma v_{t-1} + \eta \nabla_{\theta} J(\theta - \gamma v_{t-1}) \\ \theta_{t+1} &= \theta_t - v_t\end{aligned}$$

### 1.4.2 Adaptive Methods

Two main drawbacks of SGD are; tuning the learning rate is complex, and we use the same learning rate for each parameter. Adaptive methods offer solutions to these problems. They use intelligent ways to modify the learning rate that may differ from parameter to parameter, and some of them even remove the need to set the learning rate. However, they are still gradient-based algorithms with some modifications, and they do not always guarantee convergence.

#### Adagrad

In vanilla SGD and SGD with momentum, we used the same learning rate for each parameter. On the contrary, adagrad adapts the learning rates for each parameter. It performs larger updates for infrequent parameters and smaller updates for frequent parameters. To do this, it updates the learning rate at each time step  $t$  for each parameter based on their past gradients.

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{G_t + \epsilon}} \odot g_t$$

With this update rule, the learning rate is modified at each time step.  $G_t$  contains the sum of squares of the past gradients for all parameters.  $g_t$  is the gradients of all parameters at time step  $t$ , and  $\epsilon$  is the smoothing constant to avoid zero division, and it is usually set to  $10^{-8}$ .  $G_t$  is getting larger with each step since we only add positive terms that make the learning rate very small, and the algorithm cannot learn any more in advancing time steps.



## 1. Theoretical Foundations

### Adadelta

Adadelta is an extension of Adagrad, which tries to solve the decreasing learning rate problem and tries to remove the need for tuning the learning rate manually [16]. Instead of using the squares of all past gradients, Adadelta sets a moving window of gradient updates, and by doing so, it continues learning even after many iterations. It does by storing the exponentially decaying average of the squared gradients.

$$E[g^2]_t = \rho E[g^2]_{t-1} + (1 - \rho)g_t^2$$

$E[g^2]_t$  is the running average,  $\rho$  is the decay constant which is similar to momentum term (it is usually set to around 0.9 like momentum). The denominator of the update rule of adadelta is very similar to adagrad, only difference is  $G_t$  is replaced with  $E[g^2]_t$ . The term  $\sqrt{E[g^2]_t + \epsilon}$  can be rephrased as root mean squares of the previous gradients up to time  $t$ .

$$\text{RMS}[g]_t = \sqrt{E[g^2]_t + \epsilon}$$

Where  $\epsilon$  is a smoothing constant for avoiding any problem in the denominator, by using this term, we can change the update rule of Adagrad to the following:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\text{RMS}[g]_t} \odot g_t$$

For clarity, we can rephrase the update rule as follows:

$$\theta_{t+1} = \theta_t + \Delta\theta_t$$

where;

$$\Delta\theta_t = -\frac{\eta}{\text{RMS}[g]_t} \odot g_t$$

Authors of [16] pointed out that parameters updates in SGD, momentum and Adagrad doesn't match with the units of the parameters. The units relate the gradients, not the parameters. To overcome this issue they defined exponentially decaying average of parameters instead of gradients.

$$E[\Delta\theta^2]_t = \rho E[\Delta\theta^2]_{t-1} + (1 - \rho)\Delta\theta_t^2$$

## 1. Theoretical Foundations

The root means squared error of the parameters is:

$$\text{RMS}[\Delta\theta]_t = \sqrt{E[\Delta\theta^2]_t + \epsilon}$$

Since  $\text{RMS}[\Delta\theta]_t$  is unknown at time step  $t$ , it is approximated with previous time step.

$$\theta_{t+1} = \theta_t - \frac{\text{RMS}[\Delta\theta]_{t-1}}{\text{RMS}[g]_t} g_t$$

Normally this final update would remove the need of learning rate but we follow the same approach as in Pytorch implementation [17] so last term is scaled with learning rate which finally yields to update rule of Adadelta:

$$\theta_{t+1} = \theta_t - \eta \frac{\text{RMS}[\Delta\theta]_{t-1}}{\text{RMS}[g]_t} g_t$$

### **RMSProp**

RMSProp is another method that is offered to solve the decreasing learning rate problem of adagrad. Geoffrey Hinton proposed it in his neural networks for machine learning class<sup>1</sup>. It is identical to the first update rule of Adadelta that is:

$$\begin{aligned} E[g^2]_t &= \rho E[g^2]_{t-1} + (1 - \rho) g_t^2 \\ \theta_{t+1} &= \theta_t - \frac{\eta}{\sqrt{E[g^2]_t + \epsilon}} \odot g_t \end{aligned}$$

Similar to momentum constant, it is suggested to set  $\rho$  to 0.9, and  $\epsilon$  is the smoothing constant similar to previous methods' update rules.

### **ADAM**

Adam is another adaptive method that adjusts the learning rates for each parameter. It also stores an exponentially decaying average of the past gradients and past squared gradients similar to momentum. It combines the best properties of adagrad and RMSProp algorithms.

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t \\ v_t &= \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \end{aligned}$$

---

<sup>1</sup>[http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture\\_slides\\_lec6.pdf](http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf)

## 1. Theoretical Foundations

Where  $m_t$  is the estimate of the first moment of the gradients and  $v_t$  is the estimate of the second moment. However, the authors noticed that with zero initialization, these two terms are biased towards zero. Therefore they proposed bias-corrected forms of these terms to overcome this problem. It is suggested to set default values for  $\beta_1$  and  $\beta_2$  as 0.9 and 0.999.

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}$$
$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

Then the update rule is very similar to Adadelta and RMSProp that is:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t$$

With ADAM, we completed the theoretical overview of the methods that are used for the experiments. We started with the simple definition of BP, we explained how the error is propagated in BP and how gradients are calculated mathematically. We also mentioned the drawbacks of BP. Then, we moved to DFA by slightly changing the error propagation mechanism of BP, and we adapted this change to our mathematical foundations. We specified which drawbacks of BP were addressed by DFA, and we also mentioned the limitations of DFA. After that, lazy methods are summarized since they are implemented and used in experiments of previous studies. Finally, optimizers are described in general. We mentioned why they are important and how they might increase the convergence property, and lastly, we presented their update rules and drawbacks.

# 2

## Learning Problems

### Contents

---

<b>2.1</b>	<b>Parity Learning Problem . . . . .</b>	<b>19</b>
<b>2.2</b>	<b>Random Data . . . . .</b>	<b>21</b>

---

The success of neural networks spawned a great interest in the community in the fields of learnability of the various models. This involves testing different models on the same problem and observing the results. It is particularly beneficial to understand the learning dynamics of the models, which helps to find out the limitations. These studies achieved striking success in understanding the neural networks.

### 2.1 Parity Learning Problem

In [3], authors questioned how far neural networks could go beyond the linear models. They did this by focusing on parities that have a complex family of target functions. They demonstrated that this family could be approximated by a two-layer network trained with SGD, but not by lazy methods. This study brings an explanation of why neural networks' performance is better than linear methods, and it proves neural networks' learning capacities are beyond lazy methods.

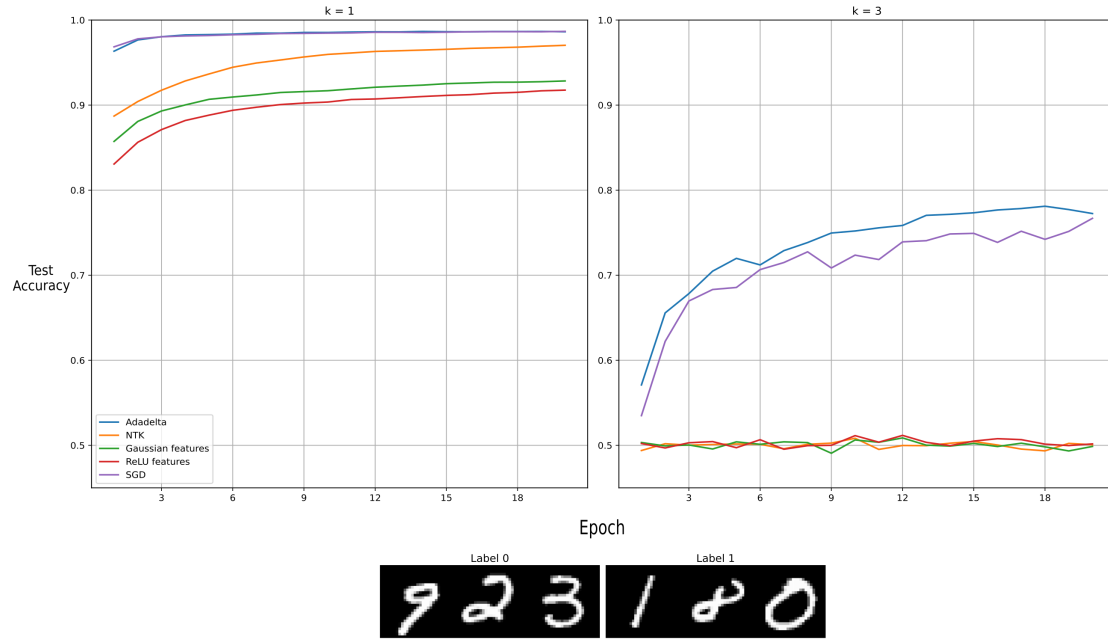
Experiments are performed on the MNIST dataset by imitating the parity problem.

## 2. Learning Problems

The task is: given the parameter  $k$  (defines the number of digits to be stacked together that is chosen uniformly from the dataset), determine if the sum of the digits is odd or even. When  $k = 1$ , it is a simplified version of the standard MNIST task to find if a digit is even or odd, and experiment results showed that all models, including the lazy ones, reached a similar performance where the neural network slightly outperformed others. On the other hand, the problem becomes more difficult for the case  $k = 3$  because models need to compute the parity of the digits' sum. In this case, there is a drastic gap between the neural network and lazy methods. Since our goal is comparing DFA and BP on this particular problem, reproducing the results from [3] is unavoidable. The same configurations are used from Pytorch [17] with minor differences. The network has only a hidden layer with 512 neurons. For the last layer, sigmoid is used as a non-linear activation function. For the hidden layer, reLU is used. BCE is preferred as a loss function, and  $10^{-3}$  is set to weight decay. We also used SGD to observe how much Adadelta improves, and for the case  $k = 3$ , we performed a simple hyper-parameter tuning process to get a decent learning rate for each method. Same values are used for the case  $k = 1$ . The hyperparameter tuning process is the following. First, we define the parameter space, later we run with all different learning rates, and we compare these runs by the average of test accuracy of the last ten epochs, and we choose the highest one. It is also crucial to mention that, at each epoch, train data is recreated to boost the available data for the models. The same is also performed for test data to have an unbiased estimation of test accuracy.

The reproduced result can be observed in figure 2.1. Similar to results from [3], all the methods succeed learning for  $k = 1$  case. However, adadelta and SGD slightly outperformed lazy methods in this setting. In the case of  $k = 3$ , adadelta and SGD almost reach 80%, but the performance of lazy methods doesn't go beyond a random guess. After having the concrete picture from the previous study, it is intriguing to see how DFA would perform with SGD and adaptive methods on this particular problem. We will investigate it in chapter 3 with other experiments.

## 2. Learning Problems



**Figure 2.1:** Reproduced MNIST parity experiment result from [3]

## 2.2 Random Data

TODO: Write how this data generated, why it's interesting and why it is suitable for testing learning dynamics of the methods.

# 3

## Citations, cross-references, and collaboration

### Contents

---

<b>3.1 Citations</b>	<b>22</b>
3.1.1 PDF output	23
3.1.2 Gitbook output	24
3.1.3 Insert references easily with the <code>citr</code> add-in	25
<b>3.2 Cross-referencing</b>	<b>25</b>
3.2.1 Section references	25
3.2.2 Figure (image and plot) references	26
3.2.3 Table references	26
3.2.4 Including page numbers	27
<b>3.3 Collaborative writing</b>	<b>28</b>
<b>3.4 Additional resources</b>	<b>29</b>

---

### 3.1 Citations

The usual way to include citations in an *R Markdown* document is to put references in a plain text file with the extension `.bib`, in **BibTeX** format.<sup>1</sup> Then reference the path to this file in `index.Rmd`'s YAML header with `bibliography: example.bib`.

---

<sup>1</sup>The bibliography can be in other formats as well, including EndNote (`.enl`) and RIS (`.ris`), see [rmarkdown.rstudio.com/authoring\\_bibliographies\\_and\\_citations](https://rmarkdown.rstudio.com/authoring_bibliographies_and_citations).

### 3. Citations and cross-refs

Most reference managers can create a .bib file with your references automatically. However, the **by far** best reference manager to use with *R Markdown* is [Zotero](#) with the [Better BibTex plug-in](#), because the `citr` plugin for RStudio (see below) can read references directly from your Zotero library!

Here is an example of an entry in a **.bib** file:

```
@article{Shea2014,  
  author = {Shea, Nicholas and Boldt, Annika},  
  journal = {Trends in Cognitive Sciences},  
  pages = {186--193},  
  title = {{Supra-personal cognitive control}},  
  volume = {18},  
  year = {2014},  
  doi = {10.1016/j.tics.2014.01.006},  
}
```

In this entry highlighted section, ‘Shea2014’ is the **citation identifier**. To default way to cite an entry in your text is with this syntax: `[@citation-identifier]`.

So I might cite some things `[Shea2014, Lottridge2012]`.

#### 3.1.1 PDF output

In PDF output, the bibliography is handled by the OxThesis LaTeX template. If you set `bib-humanities: true` in `index.Rmd`, then in-text references will be formatted as author-year; otherwise references will be shown as numbers.

If you choose author-year formatting, a number of variations on the citation syntax are useful to know:

- Put author names outside the parenthesis
  - This: `@Shea2014` says blah.
  - Becomes: **Shea2014** says blah.
- Include only the citation-year (in parenthesis)
  - This: Shea et al. says blah `[-@Shea2014]`
  - Becomes: Shea et al. says blah **[Shea2014]**



### *3. Citations and cross-refs*

- Add text and page or chapter references to the citation
  - This: [see @Shea2014, pp. 33-35; also @Wu2016, ch. 1]
  - Becomes: Blah blah [Shea2014, Wu2016].

#### **3.1.2 Gitbook output**

In gitbook output, citations are by default inserted in the Chicago author-date format.

To change the format, add `csl: some-other-style.csl` in **index.Rmd**'s YAML header. You can browse through and download styles at [zotero.org/styles](https://zotero.org/styles).

### 3. Citations and cross-refs

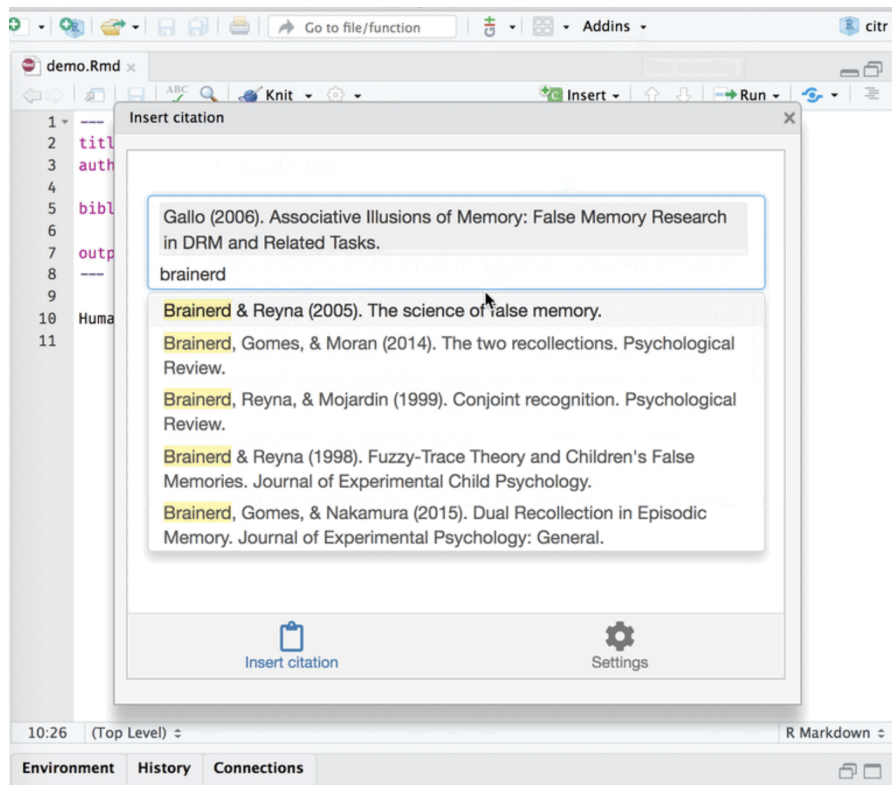


Figure 3.1: The ‘citr’ add-in

#### 3.1.3 Insert references easily with the `citr` add-in

For an easy way to insert citations, try the `citr` RStudio add-in (Figure 3.1). You can install this add-in by typing `install.packages("citr")` in the R Console.

## 3.2 Cross-referencing

We can make cross-references to **sections** within our document, as well as to **figures** (images and plots) and **tables**.

The general cross-referencing syntax is `\@ref(label)`

#### 3.2.1 Section references

Headers are automatically assigned a reference label, which is the text in lower caps separated by dashes. For example, `# My header` is automatically given the label `my-header`. So `# My header` can be referenced with `\@ref(my-section)`

### 3. Citations and cross-refs

Remember what we wrote in section 3.1?

We can also use **hyperlink syntax** and add `#` before the label, though this is only guaranteed to work properly in HTML output:

- So if we write `Remember what we wrote up in [the previous section](#citations)?`
- It becomes `Remember what we wrote up in the previous section?`

### Creating custom labels

It is a very good idea to create **custom labels** for our sections. This is because the automatically assigned labels will change when we change the titles of the sections - to avoid this, we can create the labels ourselves and leave them untouched if we change the section titles.

We create custom labels by adding `{#label}` after a header, e.g. `# My section {#my-label}`. See [our chapter title](#) for an example. That was section ??.

### 3.2.2 Figure (image and plot) references

- To refer to figures (i.e. images and plots) use the syntax `\@ref(fig:label)`
- **GOTCHA:** Figures and tables must have captions if you wish to cross-reference them.

Let's add an image:

```
knitr::include_graphics("figures/sample-content/captain.jpeg")
```

We refer to this image with `\@ref(fig:captain)`. So Figure 3.2 is [this image](#).

And in Figure ?? we saw a [cars](#) plot.

### 3.2.3 Table references

- To refer to tables use the syntax `\@ref(tab:label)`

Let's include a table:

### 3. Citations and cross-refs



**Figure 3.2:** A marvel-lous meme

**Table 3.1:** Stopping cars

speed	dist
4	2
4	10
7	4
7	22
8	16

```
knitr::kable(cars[1:5,],  
             caption="Stopping cars")
```

We refer to this table with `\@ref{tab:cars-table2}`. So Table 3.1 is [this table](#).

And in Table ?? we saw more or less [the same cars table](#).

#### 3.2.4 Including page numbers

Finally, in the PDF output we might also want to include the page number of a reference, so that it's easy to find in physical printed output. LaTeX has a command for this, which looks like this: `\pageref{fig/tab:label}` (note: curly

### 3. Citations and cross-refs

braces, not parentheses)

When we output to PDF, we can use raw LaTeX directly in our .Rmd files. So if we wanted to include the page of the cars plot we could write:

- This: `Figure \@ref(fig:cars-plot) on page \pageref(fig:cars-plot)`
- Becomes: Figure ?? on page ??

#### Include page numbers only in PDF output

A problem here is that LaTeX commands don't display in HTML output, so in the gitbook output we'd see simply "Figure ?? on page".

One way to get around this is to use inline R code to insert the text, and use an `ifelse` statement to check the output format and then insert the appropriate text.

- So this: ``r ifelse(knitr::is_latex_output(), "Figure \@ref(fig:cars-plot) on page \pageref{fig:cars-plot}", "")``
- Inserts this (check this on both PDF and gitbook): Figure ?? on page ??

Note that we need to escape the backslash with another backslash here to get the correct output.

## 3.3 Collaborative writing

Best practices for collaboration and change tracking when using R Markdown are still an open question. In the blog post [One year to dissertate](#) by Lucy D'Agostino, which I highly recommend, the author notes that she knits .Rmd files to a word document, then uses the `googledrive` R package to send this to Google Drive for comments / revisions from co-authors, then incorporates Google Drive suggestions *by hand* into the .Rmd source files. This is a bit clunky, and there are ongoing discussions among the *R Markdown* developers about what the best way is to handle collaborative writing (see [issue #1463](#) on GitHub, where [CriticMarkup](#) is among the suggestions).

### 3. Citations and cross-refs

For now, this is an open question in the community of R Markdown users. I often knit to a format that can easily be imported to Google Docs for comments, then go over suggested revisions and manually incorporate them back in to the .Rmd source files. For articles, I sometimes upload a near-final draft to [Overleaf](#), then collaboratively make final edits to the LaTeX file there. I suspect some great solution will be developed in the not-to-distant future, probably by the RStudio team.

## 3.4 Additional resources

- *R Markdown: The Definitive Guide* - <https://bookdown.org/yihui/rmarkdown/>
- *R for Data Science* - <https://r4ds.had.co.nz>

# 4

## Tables

### Contents

---

<b>4.1</b>	<b>Making LaTeX tables play nice . . . . .</b>	<b>30</b>
4.1.1	Making your table pretty . . . . .	30
4.1.2	If your table is too wide . . . . .	31
4.1.3	If your table is too long . . . . .	32
4.1.4	Max power: manually adjust the raw LaTeX output . .	41

---

## 4.1 Making LaTeX tables play nice

Dealing with tables in LaTeX can be painful. This section explains the main tricks you need to make the pain go away.

(Note: if you are looking at the ebook version, you will not see much difference in this section, as it is only relevant for PDF output!)

### 4.1.1 Making your table pretty

When you use `kable` to create tables, you will almost certainly want to set the option `booktabs = TRUE`. This makes your table look a million times better:

#### 4. Tables

```
library(knitr)
library(tidyverse)

head(mtcars) %>%
  kable(booktabs = TRUE)
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

Compare this to the default style, which looks terrible:

```
head(mtcars) %>%
  kable()
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

#### 4.1.2 If your table is too wide

You might find that your table expands into the margins of the page, like the tables above. Fix this with the `kable_styling` function from the `kableExtra` package:

```
library(kableExtra)

head(mtcars) %>%
  kable(booktabs = TRUE) %>%
  kable_styling(latex_options = "scale_down")
```

This scales down the table to fit the page width.



#### 4. Tables

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

### 4.1.3 If your table is too long

If your table is too long to fit on a single page, set `longtable = TRUE` in the `kable` function to split the table across multiple pages.

```
a_long_table <- rbind(mtcars, mtcars)

a_long_table %>%
  select(1:8) %>%
  kable(booktabs = TRUE, longtable = TRUE)
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1

#### 4. Tables

Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0
AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0
Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0
Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0
Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1
Porsche 914-2	26.0	4	120.3	91	4.43	2.140	16.70	0
Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1
Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.50	0
Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0
Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0
Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1
Mazda RX41	21.0	6	160.0	110	3.90	2.620	16.46	0
Mazda RX4 Wag1	21.0	6	160.0	110	3.90	2.875	17.02	0
Datsun 7101	22.8	4	108.0	93	3.85	2.320	18.61	1
Hornet 4 Drive1	21.4	6	258.0	110	3.08	3.215	19.44	1
Hornet Sportabout1	18.7	8	360.0	175	3.15	3.440	17.02	0
Valiant1	18.1	6	225.0	105	2.76	3.460	20.22	1
Duster 3601	14.3	8	360.0	245	3.21	3.570	15.84	0
Merc 240D1	24.4	4	146.7	62	3.69	3.190	20.00	1
Merc 2301	22.8	4	140.8	95	3.92	3.150	22.90	1
Merc 2801	19.2	6	167.6	123	3.92	3.440	18.30	1
Merc 280C1	17.8	6	167.6	123	3.92	3.440	18.90	1
Merc 450SE1	16.4	8	275.8	180	3.07	4.070	17.40	0
Merc 450SL1	17.3	8	275.8	180	3.07	3.730	17.60	0
Merc 450SLC1	15.2	8	275.8	180	3.07	3.780	18.00	0
Cadillac Fleetwood1	10.4	8	472.0	205	2.93	5.250	17.98	0
Lincoln Continental1	10.4	8	460.0	215	3.00	5.424	17.82	0
Chrysler Imperial1	14.7	8	440.0	230	3.23	5.345	17.42	0
Fiat 1281	32.4	4	78.7	66	4.08	2.200	19.47	1
Honda Civic1	30.4	4	75.7	52	4.93	1.615	18.52	1
Toyota Corolla1	33.9	4	71.1	65	4.22	1.835	19.90	1
Toyota Corona1	21.5	4	120.1	97	3.70	2.465	20.01	1
Dodge Challenger1	15.5	8	318.0	150	2.76	3.520	16.87	0
AMC Javelin1	15.2	8	304.0	150	3.15	3.435	17.30	0
Camaro Z281	13.3	8	350.0	245	3.73	3.840	15.41	0
Pontiac Firebird1	19.2	8	400.0	175	3.08	3.845	17.05	0
Fiat X1-91	27.3	4	79.0	66	4.08	1.935	18.90	1
Porsche 914-21	26.0	4	120.3	91	4.43	2.140	16.70	0
Lotus Europa1	30.4	4	95.1	113	3.77	1.513	16.90	1
Ford Pantera L1	15.8	8	351.0	264	4.22	3.170	14.50	0
Ferrari Dino1	19.7	6	145.0	175	3.62	2.770	15.50	0
Maserati Bora1	15.0	8	301.0	335	3.54	3.570	14.60	0

#### 4. Tables

Volvo 142E1	21.4	4	121.0	109	4.11	2.780	18.60	1
-------------	------	---	-------	-----	------	-------	-------	---

When you do this, you'll probably want to make the header repeat on new pages.

Do this with the `kable_styling` function from `kableExtra`:

```
a_long_table %>%
  kable(booktabs = TRUE, longtable = TRUE) %>%
  kable_styling(latex_options = "repeat_header")
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Porsche 914-2	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8

#### 4. Tables

(continued)

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2
Mazda RX41	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag1	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 7101	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive1	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout1	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant1	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 3601	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D1	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 2301	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 2801	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C1	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE1	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL1	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC1	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood1	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental1	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial1	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
Fiat 1281	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic1	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla1	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
Toyota Corona1	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Dodge Challenger1	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
AMC Javelin1	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
Camaro Z281	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
Pontiac Firebird1	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
Fiat X1-91	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Porsche 914-21	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
Lotus Europa1	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
Ford Pantera L1	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
Ferrari Dino1	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
Maserati Bora1	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
Volvo 142E1	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2

Unfortunately, we cannot use the `scale_down` option with a `longtable`. So if a `longtable` is too wide, you can either manually adjust the font size, or show the table in landscape layout. To adjust the font size, use `kableExtra`'s `font_size` option:

#### 4. Tables

```
a_long_table %>%
  kable(booktabs = TRUE, longtable = TRUE) %>%
  kable_styling(font_size = 9, latex_options = "repeat_header")
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Porsche 914-2	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2
Mazda RX41	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag1	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 7101	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive1	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout1	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant1	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 3601	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D1	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 2301	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 2801	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C1	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE1	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL1	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC1	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood1	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4

#### 4. Tables

(continued)

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Lincoln Continental1	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial1	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
Fiat 1281	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic1	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla1	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
Toyota Corona1	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Dodge Challenger1	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
AMC Javelin1	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
Camaro Z281	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
Pontiac Firebird1	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
Fiat X1-91	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Porsche 914-21	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
Lotus Europa1	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
Ford Pantera L1	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
Ferrari Dino1	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
Maserati Bora1	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
Volvo 142E1	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2

To put the table in landscape mode, use kableExtra's `landscape` function:

```
a_long_table %>%
  kable(booktabs = TRUE, longtable = TRUE) %>%
  kable_styling(latex_options = "repeat_header") %>%
  landscape()
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2

(continued)

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Porsche 914-2	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2
Mazda RX41	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag1	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 7101	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive1	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout1	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant1	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 3601	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D1	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 2301	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 2801	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C1	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE1	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL1	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC1	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood1	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental1	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial1	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4



*(continued)*

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Fiat 1281	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic1	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla1	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
Toyota Corona1	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Dodge Challenger1	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
AMC Javelin1	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
Camaro Z281	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
Pontiac Firebird1	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
Fiat X1-91	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Porsche 914-21	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
Lotus Europa1	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
Ford Pantera L1	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
Ferrari Dino1	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
Maserati Bora1	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
Volvo 142E1	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2

## 4. Tables

### 4.1.4 Max power: manually adjust the raw LaTeX output

For total flexibility, you can adjust the raw LaTeX output from `kable/kableExtra` that generates the table. Let us consider how we would do this for the example of adjusting the font size if our table is too wide: Latex has a bunch of standard commands that set an approximate font size, as shown below in Figure 4.1.

<code>\tiny</code>	Lorem ipsum
<code>\scriptsize</code>	Lorem ipsum
<code>\footnotesize</code>	Lorem ipsum
<code>\small</code>	Lorem ipsum

**Figure 4.1:** Font sizes in LaTeX

You could use these to manually adjust the font size in your longtable in two steps:

1. Wrap the longtable environment in, e.g., a `scriptsize` environment, by doing a string replacement in the output from `kable/kableExtra`
2. Add the attributes that make R Markdown understand that the table is a table (it seems R drops these when we do the string replacement)

```
our_adjusted_table <- a_long_table %>%
  kable(booktabs = TRUE, longtable = TRUE) %>%
  kable_styling(latex_options = "repeat_header") %>%
  # wrap the longtable in a tiny environment
  str_replace('\\\\begin\\{longtable\\}',
              '\\\\begin\\{scriptsize\\}\\n\\\\begin\\{longtable\\}')
  ↪ %>%
  str_replace('\\\\end\\{longtable\\}',
              '\\\\end\\{longtable\\}\\n\\\\end\\{scriptsize\\}')

#add attributes to make R Markdown treat this as a kable LaTeX
↪ table again
```

#### 4. Tables

```
our_adjusted_table %>%
  structure(format = "latex", class = "knitr_kable")
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Porsche 914-2	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2
Mazda RX41	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag1	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 7101	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive1	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout1	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant1	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 3601	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D1	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 2301	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 2801	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C1	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE1	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL1	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC1	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood1	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental1	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial1	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
Fiat 1281	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic1	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla1	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
Toyota Corona1	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Dodge Challenger1	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
AMC Javelin1	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2

#### 4. Tables

(continued)

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Camaro Z281	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
Pontiac Firebird1	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
Fiat X1-91	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Porsche 914-21	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
Lotus Europa1	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
Ford Pantera L1	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
Ferrari Dino1	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
Maserati Bora1	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
Volvo 142E1	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2

*There is grandeur in this view of life, with its several powers, having been originally breathed into a few forms or into one; and that, whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved.*

— Charles Darwin [Darwin1859]

# 5

## Customisations and extensions

### Contents

---

<b>5.1</b>	<b>Front matter</b>	<b>45</b>
5.1.1	Shorten captions shown in the list of figures (PDF)	45
5.1.2	Shorten captions shown in the list of tables (PDF)	45
<b>5.2</b>	<b>Shorten running header (PDF)</b>	<b>46</b>
<b>5.3</b>	<b>Unnumbered chapters</b>	<b>46</b>
<b>5.4</b>	<b>Beginning chapters with quotes (PDF)</b>	<b>46</b>
<b>5.5</b>	<b>Highlighting corrections (HTML &amp; PDF)</b>	<b>47</b>
5.5.1	Short, inline corrections	47
5.5.2	Blocks of added or changed material	48
5.5.3	Stopping corrections from being highlighted	48
<b>5.6</b>	<b>Apply custom font color and highlighting to text (HTML &amp; PDF)</b>	<b>48</b>
<b>5.7</b>	<b>Including another paper in your thesis - embed a PDF document</b>	<b>49</b>
<b>5.8</b>	<b>Including another paper in your thesis - R Markdown child document</b>	<b>53</b>
5.8.1	An example paper in another folder	53
5.8.2	Step 1: Include paper as a child document	54
5.8.3	Step 2: Make file paths compatible	54
5.8.4	Step 3: Make sure header levels are correct	55
5.8.5	Step 4: Make sure figure widths are correct	56
<b>5.9</b>	<b>Customizing referencing</b>	<b>57</b>
5.9.1	Using a .csl file with pandoc instead of biblatex	57
5.9.2	Customizing biblatex and adding chapter bibliographies	57
<b>5.10</b>	<b>Customizing the page headers and footers (PDF)</b>	<b>59</b>
<b>5.11</b>	<b>Diving in to the OxThesis LaTeX template (PDF)</b>	<b>60</b>
<b>5.12</b>	<b>Customising to a different university</b>	<b>60</b>
5.12.1	The minimal route	60

This chapter describes a number of additional tips and tricks as well as possible customizations to the `oxforddown` thesis.

## 5.1 Front matter

### 5.1.1 Shorten captions shown in the list of figures (PDF)

You might want your list of figures (which follows the table of contents) to have shorter (or just different) figure descriptions than the actual figure captions.

Do this using the chunk option `fig.scap` ('short caption'), for example `{r captain-image, fig.cap="A very long and descriptive (and potentially boring) caption that doesn't fit in the list of figures, but helps the reader understand what the figure communicates.", fig.scap="A concise description for the list of figures"`

### 5.1.2 Shorten captions shown in the list of tables (PDF)

You might want your list of tables (which follows the list of figures in your thesis front matter) to have shorter (or just different) table descriptions than the actual table captions.

If you are using `knitr::kable` to generate a table, you can do this with the argument `caption.short`, e.g.:

```
knitr::kable(mtcars,  
             caption = "A very long and descriptive (and  
             ↪ potentially  
             boring) caption that doesn't fit in the list of  
             ↪ figures,  
             but helps the reader understand what the figure  
             communicates.",  
             caption.short = "A concise description for the list of  
             ↪ tables")
```

## 5.2 Shorten running header (PDF)

You might want a chapter's running header (i.e. the header showing the title of the current chapter at the top of page) to be shorter (or just different) to the actual chapter title.

Do this by adding the latex command `\chaptermark{My shorter version}` after your chapter title.

For example, chapter ??'s running header is simply 'Cites and cross-refs', because it begins like this:

```
# Citations, cross-references, and collaboration {#cites-and-refs}
\chaptermark{Cites and cross-refs}
```

## 5.3 Unnumbered chapters

To make chapters unnumbered (normally only relevant to the Introduction and/or the Conclusion), follow the chapter header with `{-}`, e.g. `# Introduction {-}`.

When you do this, you must also follow the heading with these two latex commands:

```
\adjustmtc
\markboth{The Name of Your Unnumbered Chapter}{{}}
```

Otherwise the chapter's mini table of contents and the running header will show the previous chapter.

## 5.4 Beginning chapters with quotes (PDF)

The OxThesis LaTeX template lets you inject some wittiness into your thesis by including a block of type `savequote` at the beginning of chapters. To do this, use the syntax ````\block type='savequote'}`.<sup>1</sup>

---

<sup>1</sup>For more on custom block types, see the relevant section in *Authoring Books with R Markdown*.

## 5. Customisations and extensions

Add the reference for the quote with the chunk option `quote_author="my author name"`. You will also want to add the chunk option `include=knitr::is_latex_output()` so that quotes are only included in PDF output.

It's not possible to use markdown syntax inside chunk options, so if you want to e.g. italicise a book name in the reference use a 'text reference': Create a named piece of text with '(ref:label-name) My text', then point to this in the chunk option with `quote_author='(ref:label-name)'`.

### 5.5 Highlighting corrections (HTML & PDF)

For when it comes time to do corrections, you may want to highlight changes made when you submit a post-viva, corrected copy to your examiners so they can quickly verify you've completed the task. You can do so like this:

#### 5.5.1 Short, inline corrections

Highlight **short, inline corrections** by doing `[like this]{.correction}` — the text between the square brackets will then be highlighted in blue in the output.

Note that pandoc might get confused by citations and cross-references inside inline corrections. In particular, it might get confused by `"[what @Shea2014 said]{.correction}"` which becomes `[Shea2014]{.correction}` In such cases, you can use LaTeX syntax directly. The correction highlighting uses the `soul` package, so you can do like this:

- If using biblatex for references, use `"\hl{what \textcite{Shea2014} said}"`
- If using natbib for references, use `"\hl{what \cite{Shea2014} said}"`

Using raw LaTeX has the drawback of corrections then not showing up in HTML output at all, but you might only care about correction highlighting in the PDF for your examiners anyway!



## 5. Customisations and extensions

### 5.5.2 Blocks of added or changed material

Highlight entire **blocks of added or changed material** by putting them in a block of type `correction`, using the syntax ````{block type='correction'}````.<sup>2</sup> Like so:

For larger chunks, like this paragraph or indeed entire figures, you can use the `correction` block type. This environment **highlights paragraph-sized and larger blocks** with the same blue colour.

*Note that correction blocks cannot be included in word output.*

### 5.5.3 Stopping corrections from being highlighted

To turn off correction highlighting, go to the YAML header of `index.Rmd`, then:

- PDF output: set `corrections: false`
- HTML output: remove or comment out - `templates/corrections.css`

## 5.6 Apply custom font color and highlighting to text (HTML & PDF)

The lua filter that adds the functionality to highlight corrections adds two more tricks: you can apply your own choice of colour to highlight text, or change the font color. The syntax is as follows:

Here's `[some text in pink highlighting]{highlight="pink"}`  
Becomes: Here's `some text in pink highlighting`.

`[Here's some text with blue font]{color="blue"}`  
Becomes: `Here's some text with blue font`

Finally — never, ever actually do this — `[here's some text with black highlighting and yellow font]{highlight="black" color="yellow"}`  
Becomes: `here's some text with black highlighting and yellow font`

---

<sup>2</sup>In the `.tex` file for PDF output, this will put the content between `\begin{correction}` and `\end{correction}`; in gitbook output it will be put between `<div class="correction">` and `</div>`.

## 5. Customisations and extensions

The file `scripts__and__filters/colour__and__highlight.lua` implements this, if you want to fiddle around with it. It works with both PDF and HTML output.

### 5.7 Including another paper in your thesis - embed a PDF document

You may want to embed existing PDF documents into the thesis, for example if your department allows a ‘portfolio’ style thesis and you need to include an existing typeset publication as a chapter.

In gitbook output, you can simply use `knitr::include_graphics` and it should include a scrollable (and downloadable) PDF. You will probably want to set the chunk options `out.width='100%'` and `out.height='1000px'`:

```
knitr::include_graphics("figures/sample-content/pdf_embed_example/Lyngs2020_FB.pdf")
```

In LaTeX output, however, this approach can cause odd behaviour. Therefore, when you build your thesis to PDF, split the PDF into an alphanumerically sorted sequence of **single-page** PDF files (you can do this automatically with the package `pdftools`). You can then use the appropriate LaTeX command to insert them, as shown below (for brevity, in the `oxforddown` PDF sample content we’re only including two pages). *Note that the chunk option `results='asis'` must be set.* You may also want to remove margins from the PDF files, which you can do with Adobe Acrobat (paid version) and likely other software.

```
# install.packages(pdftools)
# split PDF into pages stored in
  figures/sample-content/pdf_embed_example/split/
#
↪ pdftools::pdf_split("figures/sample-content/pdf_embed_example/Lyngs2020_FB.pdf")
# output = "figures/sample-content/pdf_embed_example/split/"

# grab the pages
pages <-
↪ list.files("figures/sample-content/pdf_embed_example/split",
  full.names = TRUE)
```

## 5. Customisations and extensions

```
# set how wide you want the inserted PDFs to be:
# 1.0 is 100 per cent of the oxforddown PDF page width;
# you may want to make it a bit bigger
pdf_width <- 1.2

# for each PDF page, insert it nicely and
# end with a page break
cat(stringr::str_c("\\newpage \\begin{center}
  \\makebox[\\linewidth][c]{\\includegraphics[width=", pdf_width,
  "\\linewidth]{", pages, "}} \\end{center}"))
```

# ‘I Just Want to Hack Myself to Not Get Distracted’: Evaluating Design Interventions for Self-Control on Facebook

Ulrik Lyngs<sup>1</sup>, Kai Lukoff<sup>2</sup>, Petr Slovak<sup>3</sup>, William Seymour<sup>1</sup>, Helena Webb<sup>1</sup>,  
Marina Jirotko<sup>1</sup>, Jun Zhao<sup>1</sup>, Max Van Kleek<sup>1</sup>, Nigel Shadbolt<sup>1</sup>

<sup>1</sup>Department of Computer Science, University of Oxford, UK, {first.last}@cs.ox.ac.uk

<sup>2</sup>Human Centered Design & Engineering, University of Washington, Seattle, US, kai.l@uw.edu

<sup>3</sup>Department of Informatics, King’s College London, UK, petr.slovak@kcl.ac.uk

## ABSTRACT

Beyond being the world’s largest social network, Facebook is for many also one of its greatest sources of digital distraction. For students, problematic use has been associated with negative effects on academic achievement and general wellbeing. To understand what strategies could help users regain control, we investigated how simple interventions to the Facebook UI affect behaviour and perceived control. We assigned 58 university students to one of three interventions: goal reminders, removed newsfeed, or white background (control). We logged use for 6 weeks, applied interventions in the middle weeks, and administered fortnightly surveys. Both goal reminders and removed newsfeed helped participants stay on task and avoid distraction. However, goal reminders were often annoying, and removing the newsfeed made some fear missing out on information. Our findings point to future interventions such as controls for adjusting types and amount of available information, and flexible blocking which matches individual definitions of ‘distraction’.

## Author Keywords

Facebook; problematic use; self-control; distraction; ICT non-use; addiction; focus; interruptions

## CCS Concepts

•Human-centered computing → Empirical studies in HCI;

## INTRODUCTION

Research on ‘Problematic Facebook Use’ (PFU) has investigated correlations between Facebook use and negative effects on outcomes such as level of academic achievement [35] and subjective wellbeing [58, 57]. A cross-cutting finding is that negative outcomes are associated with difficulty at exerting self-control over use, as well as specific use patterns including viewing friends’ wide-audience broadcasts rather than receiving targeted communication from strong ties [13, 58].

Much of this work has focused on self-control over Facebook use in student populations [2, 44, 46], with media multitasking research finding that students often give in to use which provides short-term ‘guilty pleasures’ over important, but aversive academic tasks [76, 88, 60]. In the present paper, we present a mixed-methods study exploring how two interventions to Facebook — goal reminders and removing the newsfeed — affect university students’ patterns of use and perceived control over Facebook use. To triangulate self-report with objective measurement, our study combined usage logging with fortnightly surveys and post-study interviews.

We found that both interventions helped participants stay on task and use Facebook more in line with their intentions. In terms of use patterns, goal reminders led to less scrolling, fewer and shorter visits, and less time on site, whereas removing the newsfeed led to less scrolling, shorter visits, and less content ‘liked’. However, goal reminders were often experienced as annoying, and removing the newsfeed made some participants fear missing out on information. After the study, participants suggested a range of design solutions to mitigate self-control struggles on Facebook, including controls for filtering or removing the newsfeed, reminders of time spent and of use goals, and removing features that drive engagement. As an exploratory study, this work should be followed by confirmatory studies to assess whether our findings replicate, and how they may generalise beyond a student population.

## RELATED WORK

### Struggles with Facebook use

Whereas many uses of Facebook offer important benefits, such as social support, rapid spread of information, or facilitation of real-world interactions [78], a substantial amount of research has focused on negative aspects [58]. For example, studies have reported correlations between patterns of Facebook use and lower academic achievement [77, 86], low self-esteem, depression and anxiety [51], feelings of isolation and loneliness [2], and general psychological distress [15]. Such ‘Problematic Facebook Use’ (PFU) has been studied under various names (including ‘Facebook dependence’ [87] and ‘Facebook addiction’ [5]), but a recent review summarised a common definition as ‘problematic behaviour characterised by addictive-like symptoms and/or self-regulation difficulties related to Facebook use leading to negative consequences in personal and social life’ [58].

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).  
CHI ’20, April 25–30, 2020, Honolulu, HI, USA.  
© 2020 Copyright is held by the author/owner(s).  
ACM ISBN 978-1-4503-6708-0/20/04.  
<http://dx.doi.org/10.1145/3313831.3376672>

## REFERENCES

- [1] Alexander T. Adams, Jean Costa, Malte F. Jung, and Tanzeem Choudhury. 2015. Mindless Computing: Designing Technologies to Subtly Influence Behavior. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 719–730. DOI: <http://dx.doi.org/10.1145/2750858.2805843>
- [2] Sami Abdo Radman Al-Dubai, Kurubaran Ganasegeran, Mustafa Ahmed Mahdi Al-Shagga, Hematram Yadav, and John T. Arokiasamy. 2013. Adverse Health Effects and Unhealthy Behaviors among Medical Students Using Facebook. <https://www.hindawi.com/journals/tswj/2013/465161/>. (2013). DOI: <http://dx.doi.org/10.1155/2013/465161>
- [3] All Party Parliamentary Group on Social Media and Young People's Mental Health and Wellbeing. 2019. *#NewFilters to Manage the Impact of Social Media on Young People's Mental Health and Wellbeing*. Technical Report. UK Parliament.
- [4] Hunt Allcott, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. 2019. *The Welfare Effects of Social Media*. Working Paper 25514. National Bureau of Economic Research. DOI: <http://dx.doi.org/10.3386/w25514>
- [5] Cecilie Schou Andreassen, Torbjørn Torsheim, Geir Scott Brunborg, and Staale Pallesen. 2012. Development of a Facebook Addiction Scale. *Psychological Reports* 110, 2 (apr 2012), 501–517. DOI: <http://dx.doi.org/10.2466/02.09.18.PR0.110.2.501-517>
- [6] Yummy Apps. 2019. Todobook. (May 2019).
- [7] Albert Bandura. 1982. Self-efficacy mechanism in human agency. *American Psychologist* 37, 2 (1982), 122–147. DOI: <http://dx.doi.org/10.1037/0003-066x.37.2.122>
- [8] Fanni Bányaí, Ágnes Zsila, Orsolya Király, Aniko Maraz, Zsuzsanna Elekes, Mark D. Griffiths, Cecilie Schou Andreassen, and Zsolt Demetrovics. 09-Jan-2017. Problematic Social Media Use: Results from a Large-Scale Nationally Representative Adolescent Sample. *PLOS ONE* 12, 1 (09-Jan-2017), e0169839. DOI: <http://dx.doi.org/10.1371/journal.pone.0169839>
- [9] Elliot T Berkman, Cendri A Hutcherson, Jordan L Livingston, Lauren E Kahn, and Michael Inzlicht. 2017. Self-Control as Value-Based Choice. *Current Directions in Psychological Science* 26, 5 (2017), 422–428. DOI: <http://dx.doi.org/10.1177/0963721417704394>
- [10] Walter R. Boot, Daniel J. Simons, Cary Stothart, and Cassie Stutts. 2013. The Pervasive Problem with Placebos in Psychology. *Perspectives on Psychological Science* 8, 4 (jul 2013), 445–454. DOI: <http://dx.doi.org/10.1177/1745691613491271>
- [11] Amara Brook. 2011. Ecological Footprint Feedback: Motivating or Discouraging? *Social Influence* 6, 2 (April 2011), 113–128. DOI: <http://dx.doi.org/10.1080/15534510.2011.566801>
- [12] Gharad Bryan, Dean Karlan, and Scott Nelson. 2010. Commitment Devices. *Annual Review of Economics* 2, 1 (Sept. 2010), 671–698. DOI: <http://dx.doi.org/10.1146/annurev.economics.102308.124324>
- [13] Moira Burke and Robert E. Kraut. 2016. The Relationship Between Facebook Use and Well-Being Depends on Communication Type and Tie Strength. *Journal of Computer-Mediated Communication* 21, 4 (2016), 265–281. DOI: <http://dx.doi.org/10.1111/jcc4.12162>
- [14] Moira Burke, Cameron Marlow, and Thomas Lento. 2010. Social Network Activity and Social Well-Being. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 1909–1912. DOI: <http://dx.doi.org/10.1145/1753326.1753613>
- [15] Wenhong Chen and Kye-Hyoung Lee. 2013. Sharing, Liking, Commenting, and Distressed? The Pathway between Facebook Interaction and Psychological Distress. *Cyberpsychology, Behavior and Social Networking* 16, 10 (oct 2013), 728–734. DOI: <http://dx.doi.org/10.1089/cyber.2012.0272>
- [16] Justin Cheng, Moira Burke, and Elena Goetz Davis. 2019. Understanding Perceptions of Problematic Facebook Use: When People Experience Negative Life Impact and a Lack of Control. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 199:1–199:13. DOI: <http://dx.doi.org/10.1145/3290605.3300429>
- [17] Jacob Cohen. 1992. A Power Primer. *Psychological Bulletin* 112, 1 (1992), 155–159. DOI: <http://dx.doi.org/10.1037/0033-2909.112.1.155>
- [18] Anna L Cox, Sandy J J Gould, Marta E Cecchinato, Ioanna Iacovides, and Ian Renfree. 2016. Design Frictions for Mindful Interactions: The Case for Microboundaries. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 1389–1397. DOI: <http://dx.doi.org/10.1145/2851581.2892410>
- [19] Helen Creswick, Liz Dowthwaite, Ansgar Koene, Elvira Perez Vallejos, Virginia Portillo, Monica Cano, and Christopher Woodard. 2019. "... They don't really listen to people". *Journal of Information, Communication and Ethics in Society* 17, 2 (May 2019), 167–182. DOI: <http://dx.doi.org/10.1108/jices-11-2018-0090>
- [20] Angela L. Duckworth, Katherine L. Milkman, and David Laibson. 2018. Beyond Willpower: Strategies for Reducing Failures of Self-Control. *Psychological Science in the Public Interest* 19, 3 (Dec. 2018), 102–129. DOI: <http://dx.doi.org/10.1177/1529100618821893>

## 5.8 Including another paper in your thesis - R Markdown child document

Sometimes you want to include another paper you are currently writing as a chapter in your thesis. Above 5.7, we described the simplest way to do this: include the other paper as a pdf. However, in some cases you instead want to include the R Markdown source from this paper, and have it compiled within your thesis. This is a little bit more tricky, because you need to keep careful track of your file paths, but it is possible by [including the paper as a child document](#). There are four main steps:

1. Include the paper as a child document
2. Make file paths compatible with knitting the article on its own, as well as when it's include in your thesis
3. Make header levels correct
4. Make figure widths correct

### 5.8.1 An example paper in another folder

Take this simple example (files for this are in [this GitHub repository](#)):

```
|--paper_to_include
|   |--my_paper.Rmd
|   |--data
|   |   |--cat_salt.csv
|   |--figures
|   |   |--cat.jpg
|
|--thesis
```

As the chart suggests, you have another folder, **paper\_to\_include/** living in the same containing folder as your thesis folder. In the **paper\_to\_include** folder, the file **my\_paper.Rmd** is where you write the paper. In **my\_paper.Rmd**, you read in a CSV file found in the subfolder **data/cats.csv**, and also an image from the subfolder **figures/cat.jpg**.

### 5.8.2 Step 1: Include paper as a child document

In your thesis folder, create an Rmd file for the chapter where you want to include another paper. Add one or more code chunks that include R Markdown files from that paper as child documents:

```
# Including an external chapter

```{r child = "../paper_to_include/my_paper.Rmd"}
```

### 5.8.3 Step 2: Make file paths compatible

Use `parameters` to adjust the file path of images based on values you set in the YAML header of an R Markdown file. In `my_paper.Rmd`, create a parameter called `other_path` and set it to an empty string:

```
---
title: "A fabulous article in a different folder"
params:
  other_path: ""
---
```

In `my_paper.Rmd`, put this at the start of the filepath when you read in data or include images:

```
library(tidyverse)
library(knitr)

cat_data <- read_csv(str_c(params$other_path, "data/cats.csv"))
include_graphics(str_c(params$other_path, "figures/cat.jpg"))
```

Finally, in your thesis folder's `index.Rmd` file, also create the parameter `other_path`. But here, set it to where the `paper_to_include/` folder is relative to your thesis folder:

```
params:
  other_path: "../paper_to_include/"
```

## 5. Customisations and extensions

### Note on HTML output

Note that if you want to host an HTML version on your thesis online, you will need to include graphics in the content that you host online - the internet obviously won't be able to see filepaths that are just referring to stuff in another folder on your computer!

### 5.8.4 Step 3: Make sure header levels are correct

Unless the paper you want to include is also written as a book, your header levels are probably going to be off. That is, the level 1 headers (`# Some header`) you use for main sections in the other paper turns into chapter titles when included in your thesis.

To avoid this, first *increment all heading levels by one* in `paper_to_include/my_paper.Rmd` (`# Some header` -> `## Some header`). Then in `paper_to_include/` create a [lua filter](#) that decrements header levels by one: Create a text file, save it as `reduce_header_level.lua`, and give it the content below.

```
function Header(el)
  if (el.level <= 1) then
    error("I don't know how to decrease the level of h1")
  end
  el.level = el.level - 1
  return el
end
```

In the YAML header of `paper_to_include/my_paper.Rmd`, use this filter:

```
---
title: "A fabulous article in a different folder"
params:
  other_path: ""
output:
  pdf_document:
    pandoc_args: ["--lua-filter=reduce_header_level.lua"]
---
```

Now, your header levels will be correct both when you knit the paper on its own and when its included in your thesis.



## 5. Customisations and extensions

NOTE: There might be no need to use a lua filter to shift heading - it seems you could simply use `pandoc_args: ["--shift-heading-level-by=-1"]` (see <https://pandoc.org/MANUAL.html#reader-options>)

### 5.8.5 Step 4. Make sure figure widths are correct

It might be that your figure widths when knitting your paper on its own, and when including it in your thesis, need to be different. You can again use parameters to set figure widths.

Imagine you want figure width to be 80% of the page width when knitting your paper on its own, but 100% in your thesis. In `paper_to_include/my_paper.Rmd`, first add a parameter we could call `out_width` and set it to the string "80%":

```
---
title: "A fabulous article in a different folder"
params:
  other_path: ""
  out_width: "80%"
output:
  pdf_document:
    pandoc_args: ["--lua-filter=reduce_header_level.lua"]
---
```

Then, make sure use that parameter to set the output width when you include figures in `paper_to_include/my_paper.Rmd`:

```
```{r, out.width=params$out_width, fig.cap="A very funny cat"}
include_graphics(str_c(params$other_path, "figures/cat.jpg"))
```
```

Finally, create the parameter `out_width` in your thesis' `index.Rmd` file:

```
params:
  other_path: "../paper_to_include/"
  out_width: "80%"
```

Now, the output width of your figure will be 80% when knitting your paper on its own, and 100% when knitting it as child document of your thesis.

## 5.9 Customizing referencing

### 5.9.1 Using a .csl file with pandoc instead of biblatex

The `oxforddown` package uses biblatex in LaTeX for referencing. It is also possible to use pandoc for referencing by providing a .csl file in the YAML header of `index.Rmd` (likely requiring commenting out the biblatex code in `templates/template.tex`). This may be helpful for those who have a .csl file describing the referencing format for a particular journal. However, note that this approach does not support chapter bibliographies (see Section 5.9.2).

```
csl: ecology.csl
```

### 5.9.2 Customizing biblatex and adding chapter bibliographies

This section provides one example of customizing biblatex. Much of this code was combined from searches on Stack Exchange and other sources (e.g. [here](#)).

In `templates/template.tex`, one can replace the existing biblatex calls with the following to achieve referencing that looks like this:

(Charmantier and Gienapp 2014)

Charmantier, A. and P. Gienapp (2014). Climate change and timing of avian breeding and migration: evolutionary versus plastic changes. *Evolutionary Applications* 7(1):15–28. doi: 10.1111/eva.12126.

```
\usepackage[backend=biber,  
  bibencoding=utf8,  
  refsection=chapter, % referencing by chapter  
  style=authoryear,  
  firstinits=true,  
  isbn=false,  
  doi=true,  
  url=false,  
  eprint=false,  
  related=false,  
  dashed=false,
```

## 5. Customisations and extensions

```

clearlang=true,
maxcitenames=2,
mincitenames=1,
maxbibnames=10,
abbreviate=false,
minbibnames=3,
uniquelist=minyear,
sortcites=true,
date=year
]{biblatex}
\AtEveryBibitem{%
  \clearlist{language}%
  \clearfield{note}
}

\DeclareFieldFormat{titlecase}{\MakeTitleCase{#1}}

\newrobustcmd{\MakeTitleCase}[1]{%
  ↪ \ifthenelse{\ifcurrentfield{booktitle}\OR\ifcurrentfield{booksubtitle}%
    \OR\ifcurrentfield{maintitle}\OR\ifcurrentfield{mainsubtitle}%
    ↪ \OR\ifcurrentfield{journaltitle}\OR\ifcurrentfield{journalsubtitle}%
    ↪ \OR\ifcurrentfield{issuetitle}\OR\ifcurrentfield{issuesubtitle}%
    ↪ \OR\ifentrytype{book}\OR\ifentrytype{mvbook}\OR\ifentrytype{bookinbook}%
    \OR\ifentrytype{booklet}\OR\ifentrytype{suppbook}%
    \OR\ifentrytype{collection}\OR\ifentrytype{mvcollection}%
    \OR\ifentrytype{suppcollection}\OR\ifentrytype{manual}%
    \OR\ifentrytype{periodical}\OR\ifentrytype{suppperiodical}%
    \OR\ifentrytype{proceedings}\OR\ifentrytype{mvproceedings}%
    \OR\ifentrytype{reference}\OR\ifentrytype{mvreference}%
    \OR\ifentrytype{report}\OR\ifentrytype{thesis}}
    {#1}
    {\MakeSentenceCase{#1}}}

% \renewbibmacro{in:}{}
% suppress "in" for articles
%
\renewbibmacro{in:}{%
  \ifentrytype{article}{}{\printtext{\bibstring{in}\intitlepunct}}}
%-- no "quotes" around titles of chapters/article titles
\DeclareFieldFormat{article, inbook, incollection, inproceedings,
  ↪ misc, thesis, unpublished}

```

## 5. Customisations and extensions

```
{title}{#1}
%-- no punctuation after volume
\DeclareFieldFormat[article]
{volume}{{#1}}
%-- puts number/issue between brackets
\DeclareFieldFormat[article, inbook, incollection, inproceedings,
↪ misc, thesis, unpublished]
{number}{\mkbibparens{#1}}
%-- and then for articles directly the pages w/o any "pages" or
↪ "pp."
\DeclareFieldFormat[article]
{pages}{#1}
%-- for some types replace "pages" by "p."
\DeclareFieldFormat[inproceedings, incollection, inbook]
{pages}{p. #1}
%-- format 16(4):224--225 for articles
\renewbibmacro*{volume+number+eid}{
  \printfield{volume}%
  \printfield{number}%
  \printunit{\addcolon}
}
```

If you would like chapter bibliographies, in addition insert the following code at the end of each chapter, and comment out the entire REFERENCES section at the end of template.tex.

```
\printbibliography[segment=\therefsection,heading=subbibliography]
```

## 5.10 Customizing the page headers and footers (PDF)

This can now be done directly in **index.Rmd**'s YAML header. If you are a LaTeX expert and need further customisation that what's currently provided, you can tweak the relevant sections of **templates/template.tex** - the relevant code is beneath the line that begins `\usepackage{fancyhdr}`.

## 5.11 Diving in to the OxThesis LaTeX template (PDF)

For LaTeX minded people, you can read through `templates/template.tex` to see which additional customisation options are available as well as `templates/ociamthesis.cls` which supplies the base class. For example, `template.tex` provides an option for master's degree submissions, which changes identifying information to candidate number and includes a word count. At the time of writing, you must set this directly in `template.tex` rather than from the YAML header in `index.Rmd`.

## 5.12 Customising to a different university

### 5.12.1 The minimal route

If the front matter in the OxThesis LaTeX template is suitable to your university, customising `oxforddown` to your needs could be as simple as putting the name of your institution and the path to your university's logo in `index.Rmd`:

```
university: University of You
university-logo: figures/your-logo-here.pdf
```

### 5.12.2 Replacing the entire title page with your required content

If you have a `.tex` file with some required front matter from your university that you want to replace the OxThesis template's title page altogether, you can provide a filepath to this file in `index.Rmd`. `oxforddown`'s sample content includes an example of this — if you use the YAML below, your front matter will look like this:

```
alternative-title-page:
↪ front-and-back-matter/alt-title-page-example.tex
```

5. Customisations and extensions

|                                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>Title of your Thesis</p> <p>John Doe</p>                                                                                                                                      | <p><b>Thesis committee</b></p> <p><b>Promotor:</b><br/>Prof.dr. J. Smith<br/>Professor of Geo-information Science and Remote Sensing<br/>Wageningen University</p> <p><b>Co-promotors:</b><br/>Dr. Name of co-promotor<br/>Assistant Professor, Laboratory of Geo-information Science and Remote Sensing<br/>Wageningen University</p> <p><b>Other members:</b><br/>Prof.dr. Jury member 1, Wageningen University<br/>Prof.dr. Jury member 2, Affiliation<br/>Prof.dr. Jury member 3, Affiliation<br/>Prof.dr. Jury member 4, Affiliation</p> <p>This research was conducted under the auspices of the C.T. de Wit Graduate School of Production Ecology &amp; Resource Conservation (PERC)</p> | <p>Title of your thesis</p> <p>John Doe</p> <p><b>Thesis</b><br/>submitted in fulfillment of the requirements for the degree of doctor at<br/>Wageningen University<br/>by the authority of the Rector Magnificus<br/>Prof. Dr A.P.J. Mol,<br/>in the presence of the<br/>Thesis Committee appointed by the Academic Board<br/>to be defended in public<br/>on Date of your defense<br/>at 4 p.m. in the Aula</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| <p>John Doe<br/>Title of your thesis<br/>77 pages<br/>PhD thesis, Wageningen University, Wageningen, NL (2015)<br/>With references, with summary in English<br/>ISBN XXX-YYY</p> | <p>For Yihai Xie</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | <p>Acknowledgements</p> <p>This is where you will normally thank your advisor, colleagues, family and friends, as well as funding and institutional support. In our case, we will give our thanks to the people who developed the ideas and tools that allow us to push open science a little step forward by writing plain-text, transparent, and reproducible theses in R Markdown.</p> <p>We must be grateful to John Gruber for inventing the original version of Markdown, to John MacFarlane for creating Pandoc (<a href="http://pandoc.org">http://pandoc.org</a>) which converts Markdown to a large number of output formats, and to Yihai Xie for creating knitr which introduced R Markdown as a way of embedding code in Markdown documents, and bookdown which added tools for technical and longer-form writing.</p> <p>Special thanks to Chetani Ismay, who created the <i>thesisdown</i> package that helped many a PhD student write their theses in R Markdown. And a very special thanks to John MacFarlane, whose adoption of Stan Evans' adaptation of Keith Gilmer's original <i>marble</i> template for writing an Oxford University DPhil thesis in L<sup>A</sup>T<sub>E</sub>X provided the template that I in turn adapted for R Markdown.</p> <p>Finally, profuse thanks to JJ Allaire, the founder and CEO of RStudio, and Hadley Wickham, the mastermind of the tidyverse without whom we'd all just given up and done data science in Python instead. Thanks for making data science easier, more accessible, and more fun for us all.</p> <p>Ulrik Lengen<br/>Linacre College, Oxford<br/>2 December 2018</p> |

# 6

## Troubleshooting

This chapter describes common errors you may run into, and how to fix them.

### 6.1 Error: Failed to build the bibliography via biber

This can happen if you've had a failed build, perhaps in relation to RStudio shutting down abruptly.

Try doing this:

1. type `make clean-knits` in the terminal tab (or run `file.remove(list.files(pattern = "*(log|mtc|maf|aux|bbl|blg|xml)"))` in the R console) to clean up files generated by LaTeX during a build
2. restart your computer

If this does not solve the problem, try using the [natbib](#) LaTeX package instead of [biblatex](#) for handling references. To do this, go to **index.Rmd** and

1. set `use-biblatex: false` and `use-natbib: true`
2. set `citation_package: natbib` under

## 6. *Troubleshooting*

```
output:  
  bookdown::pdf_book:  
    citation_package: natbib
```



*Alles Gescheite ist schon gedacht worden.  
Man muss nur versuchen, es noch einmal zu denken.  
All intelligent thoughts have already been thought;  
what is necessary is only to try to think them again.*

— Johann Wolfgang von Goethe  
[von\_goethe\_wilhelm\_1829]

## Conclusion

If we don't want Conclusion to have a chapter number next to it, we can add the `{-}` attribute.

## More info

And here's some other random info: the first paragraph after a chapter title or section head *shouldn't be* indented, because indents are to tell the reader that you're starting a new paragraph. Since that's obvious after a chapter or section title, proper typesetting doesn't add an indent there.

This paragraph, by contrast, *will* be indented as it should because it is not the first one after the 'More info' heading. All hail LaTeX. (If you're reading the HTML version, you won't see any indentation - have a look at the PDF version to understand what in the earth this section is babbling on about).

# Appendices



## Backpropagation with Binary Cross-Entropy

Let's consider a simple binary classification task. It is common to use a network with a single logistic output with the binary cross-entropy loss function and for the sake of simplicity, let's assume that there is only one hidden layer.

$$BCE = - \sum_{i=1}^{nout} (y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i))$$

Where  $y$  is the ground truth and  $\hat{y}$  is the output of the network. After having the loss function, let's continue with the forward pass.

$$a_k = h_{k-1}w_k + b_k$$

$$h_k = f(a_k)$$

Where,  $w_k$  is the weight,  $b_k$  is the bias term,  $h_k$  is the output of the layer (which means that  $h_0 = X$  and  $h_2 = \hat{y}$ ) and  $f$  is the non linear function. Please note that for last layer logistic function is used whereas for hidden layer reLU is used as non linear functions.

We can compute the derivative of the weights by using the chain rule.

$$\frac{\partial BCE}{\partial w_2} = \frac{\partial BCE}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial a_2} \frac{\partial a_2}{\partial w_2}$$

### A. Backpropagation with Binary Cross-Entropy

Computing each factor in the term, we have:

$$\begin{aligned}\frac{\partial BCE}{\partial \hat{y}} &= \frac{-y}{\hat{y}} + \frac{1-y}{1-\hat{y}} \\ &= \frac{\hat{y}-y}{\hat{y}(1-\hat{y})} \\ \frac{\partial \hat{y}}{\partial a_2} &= \hat{y}(1-\hat{y}) \\ \frac{\partial a_2}{\partial w_2} &= h_1^T\end{aligned}$$

Which gives us:

$$\frac{\partial BCE}{\partial w_2} = h_1^T (\hat{y} - y)$$

Derivative of the  $w_1$  concerning loss function can be calculated as the following:

$$\frac{\partial BCE}{\partial w_1} = \frac{\partial BCE}{\partial h_1} \frac{\partial h_1}{\partial a_1} \frac{\partial a_1}{\partial w_1}$$

Compute each factor in the term again, we have:

$$\begin{aligned}\frac{\partial BCE}{\partial h_1} &= \frac{\partial BCE}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial a_2} \frac{\partial a_2}{\partial h_1} \\ &= (\hat{y} - y) w_2^T \\ \frac{\partial h_1}{\partial a_1} &= f'(a_1) \\ \frac{\partial a_1}{\partial h_1} &= X^T\end{aligned}$$

Which gives us:

$$\frac{\partial BCE}{\partial w_1} = (X)^T (\hat{y} - y) (w_2^T) \odot f'(a_1)$$

Where  $\odot$  is element-wise multiplication, similarly, bias terms can be calculated by following:

$$\begin{aligned}\frac{\partial BCE}{\partial b_2} &= \frac{\partial BCE}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial a_2} \frac{\partial a_2}{\partial b_2} \\ &= (\hat{y} - y) \\ \frac{\partial BCE}{\partial b_1} &= \frac{\partial BCE}{\partial h_1} \frac{\partial h_1}{\partial a_1} \frac{\partial a_1}{\partial b_1} \\ &= (\hat{y} - y) (w_2^T) \odot f'(a_1)\end{aligned}$$

### *A. Backpropagation with Binary Cross-Entropy*

After having all these results, we can update the parameters (weights and biases) using gradient descent and its variants.

B

Reproducibility

## References

- [1] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. “Learning Representations by Back-propagating Errors”. In: *Nature* 323.6088 (1986), pp. 533–536. DOI: [10.1038/323533a0](https://doi.org/10.1038/323533a0). URL: <http://www.nature.com/articles/323533a0>.
- [2] Arild Nøkland. *Direct Feedback Alignment Provides Learning in Deep Neural Networks*. 2016. arXiv: [1609.01596](https://arxiv.org/abs/1609.01596) [stat.ML].
- [3] Amit Daniely and Eran Malach. “Learning Parities with Neural Networks”. In: *CoRR* abs/2002.07400 (2020). arXiv: [2002.07400](https://arxiv.org/abs/2002.07400). URL: <https://arxiv.org/abs/2002.07400>.
- [4] Yoshua Bengio et al. *Towards Biologically Plausible Deep Learning*. 2016. arXiv: [1502.04156](https://arxiv.org/abs/1502.04156) [cs.LG].
- [5] Dong-Hyun Lee et al. *Difference Target Propagation*. 2015. arXiv: [1412.7525](https://arxiv.org/abs/1412.7525) [cs.LG].
- [6] Wan-Duo Kurt Ma, J. P. Lewis, and W. Bastiaan Kleijn. *The HSIC Bottleneck: Deep Learning without Back-Propagation*. 2019. arXiv: [1908.01580](https://arxiv.org/abs/1908.01580) [cs.LG].
- [7] Timothy P. Lillicrap et al. *Random feedback weights support learning in deep neural networks*. 2014. arXiv: [1411.0247](https://arxiv.org/abs/1411.0247) [q-bio.NC].
- [8] Julien Launay et al. *Direct Feedback Alignment Scales to Modern Deep Learning Tasks and Architectures*. 2020. arXiv: [2006.12878](https://arxiv.org/abs/2006.12878) [stat.ML].
- [9] Maria Refinetti et al. *Align, then memorise: the dynamics of learning with feedback alignment*. 2021. arXiv: [2011.12428](https://arxiv.org/abs/2011.12428) [stat.ML].
- [10] Julien Launay, Iacopo Poli, and Florent Krzakala. *Principled Training of Neural Networks with Direct Feedback Alignment*. 2019. arXiv: [1906.04554](https://arxiv.org/abs/1906.04554) [stat.ML].
- [11] Lenaïc Chizat, Edouard Oyallon, and Francis Bach. *On Lazy Training in Differentiable Programming*. 2020. arXiv: [1812.07956](https://arxiv.org/abs/1812.07956) [math.OA].
- [12] Jaehoon Lee et al. *Deep Neural Networks as Gaussian Processes*. 2018. arXiv: [1711.00165](https://arxiv.org/abs/1711.00165) [stat.ML].
- [13] Pedro Domingos. *Every Model Learned by Gradient Descent Is Approximately a Kernel Machine*. 2020. arXiv: [2012.00152](https://arxiv.org/abs/2012.00152) [cs.LG].
- [14] Arthur Jacot, Franck Gabriel, and Clément Hongler. “Neural Tangent Kernel: Convergence and Generalization in Neural Networks”. In: *CoRR* abs/1806.07572 (2018). arXiv: [1806.07572](https://arxiv.org/abs/1806.07572). URL: <http://arxiv.org/abs/1806.07572>.
- [15] Sebastian Ruder. “An overview of gradient descent optimization algorithms”. In: *CoRR* abs/1609.04747 (2016). arXiv: [1609.04747](https://arxiv.org/abs/1609.04747). URL: <http://arxiv.org/abs/1609.04747>.

## References

- [16] Matthew D. Zeiler. *ADADELTA: An Adaptive Learning Rate Method*. 2012. arXiv: [1212.5701](https://arxiv.org/abs/1212.5701) [cs.LG].
- [17] Adam Paszke et al. “PyTorch: An Imperative Style, High-Performance Deep Learning Library”. In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach et al. Curran Associates, Inc., 2019, pp. 8024–8035. URL: <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>.