

# Suggestions for the “Dissertation preparation” document.

Ivo Timóteo

11<sup>th</sup> February, 2016

---

Next time make the .tex compile without errors so that I can suggest inline.

Also, careful with commas and sentence structure. It’s better to have multiple shorter, clear sentences than nested monsters requiring the reader to keep a stack of depth 4. And don’t separate conjunctions with commas, please!

Furthermore, I am not correcting everything because I am too lazy to type. Make sure to give another very good passage over the text afterwards.

## List of figures

- Use short captions on the figures otherwise LoF looks awful.

```
\caption[Short version for LoF]{Full caption}
```

## Introduction

- **REPLACE** [first sentence]  
**BY** Brief overview of the report:
- **REPLACE** the objectives of it  
**BY** its objectives
- **REPLACE** help me achieve it  
**BY** help me achieve them
- **SPELLING** theoreical
- **REMOVE** for the unsuspicious reader that is not versed in that area of knowledge
- **REMOVE** are presented, in consecutive order.
- **REPLACE** I will use in the final solution, that will be outlined  
**BY** will be used in the final solution, which is outlined
- **REMOVE** exhaustively

- **REPLACE** in tasks  
BY as tasks
- **REPLACE** Chart  
BY chart
- **REPLACE** utilizing during the course of my work  
BY using in my work
- **REPLACE** that I have already done to prepare this thesis  
BY already concluded

## Background

- **REMOVE** The
- **REPLACE** learning structures  
BY models
- **REMOVE** , itself,
- **REPLACE** its  
BY their
- **SPELLING** beign
- **REMOVE** that extend
- **WTF** faltacitação
- **COMMENT** Not really clear at all from the short explanation why LSTM improves over HMM and RNN which is what your sentence kind of makes us expect...

## Motivation

- **REMOVE** [comma after realization]
- **REMOVE** , hitherto,
- **REPLACE** still needs to be  
BY is
- **REMOVE** possible
- **REMOVE** [comma after automatic]
- **REPLACE** , and saving  
BY thus saving
- **WTF** performance room

## Objectives

- **REPLACE** done  
BY presented
- **REMOVE** , to make justice to the name thesis,

## Theoretical background

- **COMMENT** Not all learning is for classification; not all learning wants to extract features (e.g., we might just want a simple blackbox predictor); not all learning is supervised. Explicit vector representations have commas separating different components.
- **REMOVE** diametrically
- **COMMENT** refer that the sum of squares error is common because it's what appears under the assumption of Gaussian noise.
- **COMMENT** “try to extract higher level information from data”. Can you explain? Weren't you just looking for  $y_i$ ? What is a higher level prediction?

## Artificial Neural Networks

- **COMMENT** “try to mimic the Human Brain” – I don't think this is actually accepted at all; no one I know believes this to be true. Also, why the capitalization? Also, right after that you claim to know the underlying mathematical function of a human neuron. Finally, you present a linear model as being equivalent to a function which is incredibly reductive (can't you have a function that is a non-linear model of its inputs?).
- **COMMENT**  $b_0$  is usually called  $w_0$  and you define  $x = [1, x_1, \dots, x_n]$  giving you  $y = f(\mathbf{w}^T \mathbf{x})$ . Also, you have  $N$  as number of data points and  $n$  as number of features of the input which can be confusing.
- **COMMENT** Step functions are usually between 0 and 1. On the plot it would be nicer if every function had the same range.
- **COMMENT** Why are layers important?
- **COMMENT** When discussing the equations (2.3) to (2.6) it would be nice to have the schematics of at least three chained neurons.

## Recurrent Neural Networks

- **REMOVE** , essentially,

- **REPLACE** especially  
**BY** commonly
- **COMMENT** “but no information about the correlation between themselves and the data points that preceded them did not influence the training step” – this is a double negative which I don’t think you wanted.
- **REPLACE** They were regarded as if no temporal relationship existed, and therefore each data point is conditionally independent of any other.  
**BY** Temporal relationship is disregarded and each data point considered conditionally independent of any other.
- **REPLACE** truth  
**BY** true
- **REPLACE** , in which it is actually completely false,  
**BY** such
- **SPELLING** Simulatneous
- **COMMENT** Stop the weird unnecessary capitalization of stuff :)

## LSTMN

- **REMOVE** basis
- **REPLACE** its formulation  
**BY** it
- **REPLACE** structure [2] last year  
**BY** model last year [2]
- **SPELLING** presenation
- **REPLACE** as well as the work  
**BY** as well as of the work
- **COMMENT** “Hadamard elementwise matrix multiplication” is **very** redundant. Why is the activation function of the forget gate always sigmoid? Btw, even though sigmoid is usually seen as the logistic function it actually refers to the entire family of S-shaped functions.
- **REMOVE** [commas around “can be any”]
- **REPLACE** As for the update rule is concerned, we have  
**BY** The update rule is
- **SPELLING** hipothetical
- **COMMENT** how do you assess whether the behaviour of the weight is appropriate or not?

### **Proposed Solution**

- **REPLACE** work, their  
BY work and their
- **SPELLING** acheived

### **State of the art**

- **COMMENT** “bleeding edge” is a bit too informal. Also, too many adjectives in 3.1; staggering, incredibly, astonishing, astounding, etc
- **SPELLING** Tthere
- **SPELLING** restric

### **Work plan**

- **COMMENT** I don’t know what to say about the first paragraph...
- **SPELLING** subtaskswwhich
- **SPELLING** throgh
- **SPELLING** asses
- **SPELLING** [T3.1 is badly typed]