# Machine Learning for Social Scientists

Jorge Cimentada

2020-03-02

# Contents

# Preface

Notes, content and exercises for the RECSM 2020 course Machine Learning for Social Scientists. These are intended to introduce social scientists to concepts in machine learning using traditional social science examples and datasets. Currently, it is not intended to be a book but rather supporting material for the course. Perhaps it evolves enough to be a book some day.

# Chapter 1

# Machine Learning for Social Scientists

Machine Learning practitioners and Social Scientists share many things in common. These shared traits are mostly related to the transformation, analysis and evaluation of statistical models. In fact, when many of my fellow social scientists take any introductory course on machine learning, I often hear that many of the things they get taught are very common in traditional statistics classes, if not the same. This is good news! This means that you already have a foot inside the field without even knowing it. Machine Learning practitioners use many of the same statistical model we use and also many of transformation techniques that we use. However, there are important differences on how we analyze data and how we answer our questions. In this chapter I will elaborate on how machine learning practitioners have developed strategies different from social scientists for analyzing their data, how their analysis workflow compares to ours and finally, a tour around their way of thinking, which has evolved to be very different from ours.

I hope that by understanding the strategies and techniques that machine learning practitioners use, social scientists would expand their analysis toolbox, allowing us to complement their way of thinking with our strong research design skills and modelling techniques.

## 1.1 A different way of thinking

The first question we want to ask ourselves is, what is machine learning? Machine Learning bears indeed a fancy name which brings to mind thoughts related to artificial intelligence and robots. However, as you'll see throughout the course, most terms and models used in machine learning are actually what