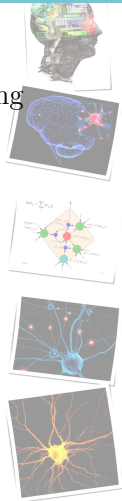


<p>INTRODUCTION</p> <div>  </div> <p>A quick introduction to machine learning</p> <p>Spyros Samothrakis Senior Lecturer, IADS University of Essex MiSoC</p> <p>June 22, 2022</p> <p>1 / 8</p>	<p>INTRODUCTION</p> <h2>WELCOME/COURSE CONTENTS</h2> <ul style="list-style-type: none"> ▶ What will this course cover? <ul style="list-style-type: none"> ▶ Day 1: An intro to machine learning (ML) ▶ Day 1: ML labs ▶ Day 2: An intro to causal inference ▶ Day 2: ML and causal inference labs ▶ Textbooks? <ul style="list-style-type: none"> ▶ Mitchell, T. M. (1997). Machine learning.¹ ▶ Bishop, C. M. (2006). Pattern recognition and machine learning. springer.² ▶ Wasserman, L. (2013). All of statistics: a concise course in statistical inference. Springer Science & Business Media.³ <p>¹http://www.cs.cmu.edu/~tom/mlbook.html ²https://www.microsoft.com/en-us/research/publication/pattern-recognition-machine-learning/ ³http://www.stat.cmu.edu/~larry/all-of-statistics/index.html</p> <p>2 / 8</p>
<p>INTRODUCTION</p> <h2>BETTER SCIENCE THROUGH DATA</h2> <p>Hey, Tony, Stewart Tansley, and Kristin M. Tolle. “Jim Gray on eScience: a transformed scientific method.” (2009).⁴</p> <ul style="list-style-type: none"> ▶ Thousand years ago: empirical branch <ul style="list-style-type: none"> ▶ You observed stuff and you wrote down about it ▶ Last few hundred years: theoretical branch <ul style="list-style-type: none"> ▶ Equations of gravity, equations of electromagnetism ▶ Last few decades: computational branch <ul style="list-style-type: none"> ▶ Modelling at the micro level, observing at the macro level ▶ Today: data exploration <ul style="list-style-type: none"> ▶ Let machines create models using vast amounts of data <p>⁴http://languagelog.ldc.upenn.edu/myl/JimGrayOnE-Science.pdf</p> <p>3 / 8</p>	<p>INTRODUCTION</p> <h2>BETTER BUSINESS THROUGH DATA</h2> <ul style="list-style-type: none"> ▶ There was a report by Mckinsey <p>Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.⁵</p> <ul style="list-style-type: none"> ▶ Urges everyone to monetise “Big Data” ▶ Use the data provided within your organisation to gain insights ▶ Has some numbers as to how much this is worth ▶ Proposes a number of methods, most of them associated with machine learning and databases <p>⁵http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation</p> <p>4 / 8</p>
<p>INTRODUCTION</p> <h2>WHY IS IT POPULAR NOW?</h2> <ul style="list-style-type: none"> ▶ Algorithms + data + tools ▶ Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). Statistical science, 16(3), 199-231.⁶ ▶ Anderson, P. W. (1972). More is different. Science, 177(4047), 393-396.⁷ ▶ Pedregosa, et.al. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, 2825-2830.⁸ <p>⁶http://projecteuclid.org/download/pdf_1/euclid.ss/1009213726%20 ⁷https://www.tkm.kit.edu/downloads/TKM1_2011_more_is_different_PWA.pdf ⁸https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf</p> <p>5 / 8</p>	<p>INTRODUCTION</p> <h2>SO THIS COURSE COVERS TOOLS</h2> <ul style="list-style-type: none"> ▶ ML theory <ul style="list-style-type: none"> ▶ <i>Supervised learning, (non)-linear regression</i> ▶ Understanding basic modelling ▶ Confirming your model is sane ▶ Tuning your model ▶ All within a very applied setting ▶ Tools <ul style="list-style-type: none"> ▶ Numpy ▶ Scikit-learn <p>6 / 8</p>

WHAT IS SUPERVISED LEARNING?

- ▶ Imagine someone gives you a group of smokers
 - ▶ And asks the question – what is there life expectancy?
- ▶ **Completely made up imaginary data**

SOME ABSTRACTION

- ▶ We are given inputs $x_0, x_1 \dots x_n$ and we are looking to predict y
- ▶ Let's plot!