AB Test Crash Course

1

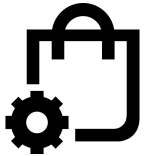
Course Introduction

Who is this course directed to:

Data Analysts

Data Scientists

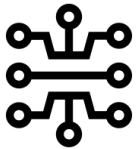
Data Engineers



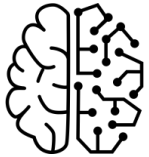
Product

What to expect from this crash course:

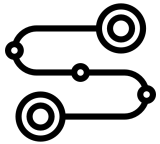
Not a lot of books exist yet for some of the topics in this course.



State-of-the-art techniques, prepared to be understood and used right away (plug and play)



Machine Learning plays an important role, and it's more straightforward than you may think.



This course is a journey. Don't start a lesson without understanding the previous one.

There will be math, but kept to a minimum and always done in Python (API approach)

Ask away. There is no such thing as a stupid question.

What AB Test can help you with:

- **Randomized Trials**
  Randomly assign participants into experimental and control groups. If done properly, the only difference between groups will be in the desired variable. This will mean that correlation will imply causation.

- **Hypothesis Testing**
  Assess plausibility of a hypothesis using population data.

- **Metric Definition**
  Better knowledge of business metrics through iteration

- **Quantitative data**
  AB test deals with estimations.

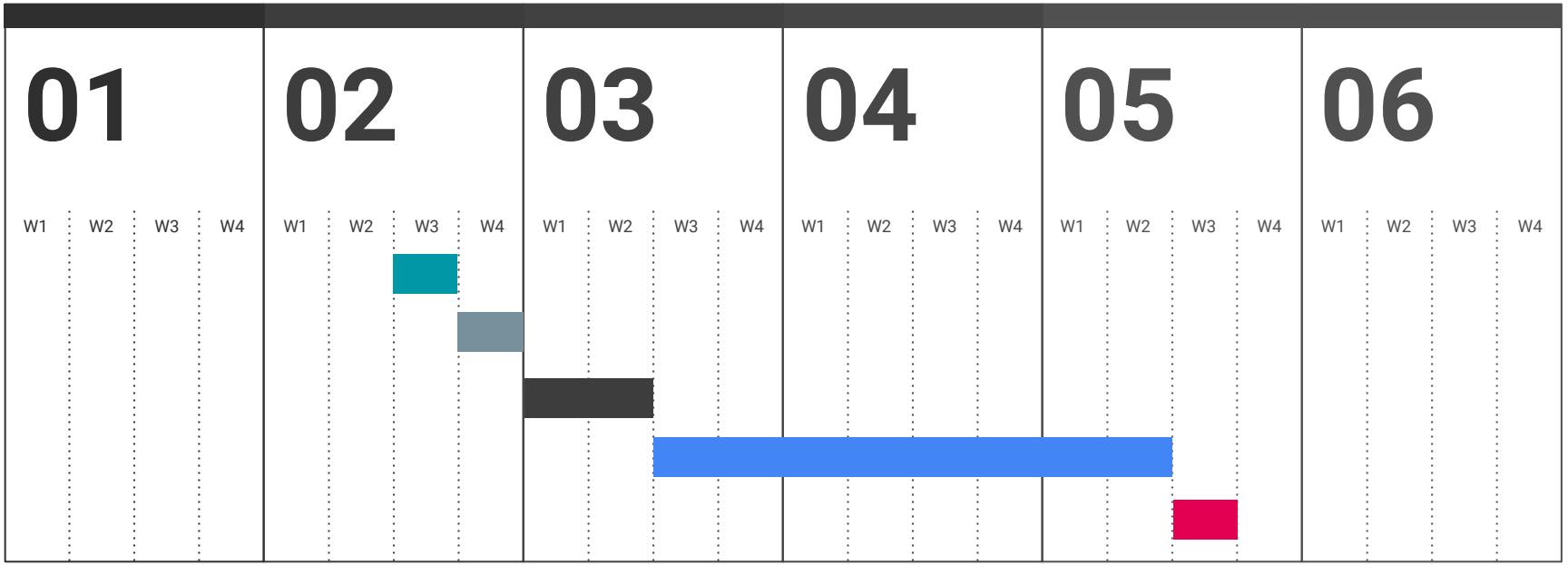
What AB Testing isn't good for:

- **New Experiences**
  Change aversion, Novelty Effect.

- **Tell you if something is missing**

- **Give you deep qualitative data**
  UX Research, Focus groups, human evaluation

AB Test process



Kick-off meet: Product, marketing or other team interested in the experiment. What do we want to achieve?

Metrics Definition & One-Pager: Robustness, importance, hypothesis. Design and duration of split.

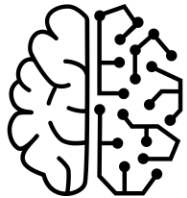
Split: Tech team

Experiment

Data gathering and Analysis

Why Machine Learning?

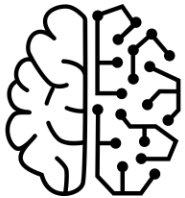
Bayesian statistics, Bayesian Machine Learning



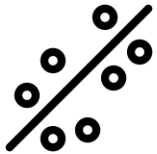
ML: A model learns its parameters from data

Why Machine Learning?

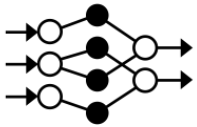
Bayesian statistics, Bayesian Machine Learning



ML: A model learns its parameters from data



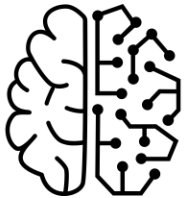
Linear Regression



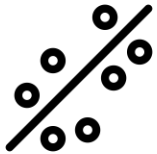
Neural Networks

Why Machine Learning?

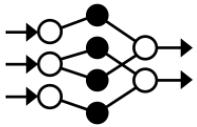
Bayesian statistics, Bayesian Machine Learning



ML: A model learns its parameters from data



Linear Regression



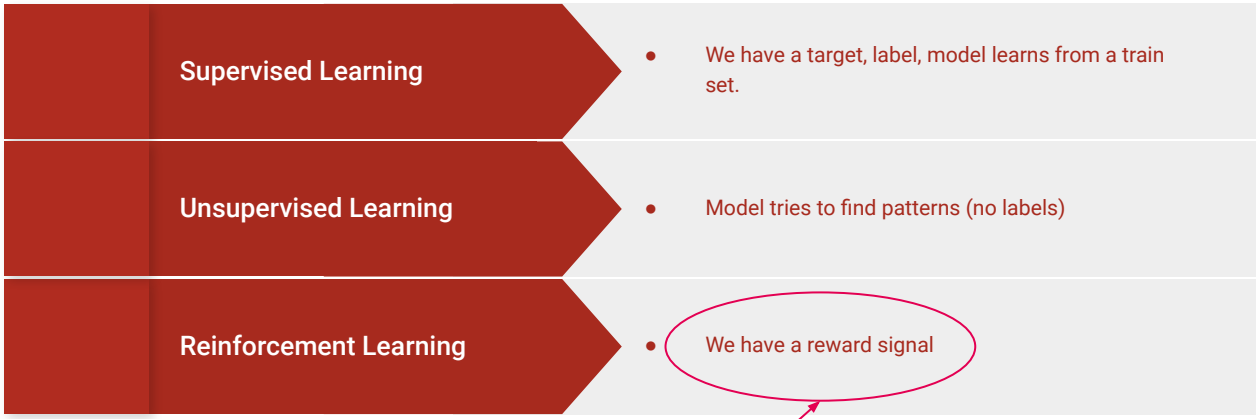
Neural Networks

The distinction between statistics and Machine Learning is not clear.
Many things we may think are ML are simply statistics...
and the other way around!

The process (math) is the same, regardless of what name we give it.

Why Machine Learning?

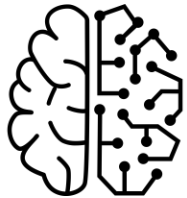
Bayesian statistics, Bayesian Machine Learning



Maximize clicks, revenue, purchases, etc

Why Machine Learning?

Bayesian statistics, Bayesian Machine Learning



Machine Learning is building a model where its parameters are learned from the data we have.

Online learning **is a type of ML where learning happens in real time.**

Data is ingested one sample at at time and parameters are updated.

The algorithm becomes *smarter* by each collected sample.

Why Machine Learning?

Bayesian statistics, Bayesian Machine Learning

This course = *Application of AB Testing + Underlayer of Bayesian Machine Learning*

Goals:

AB Testing and easy, simple reporting of results

-

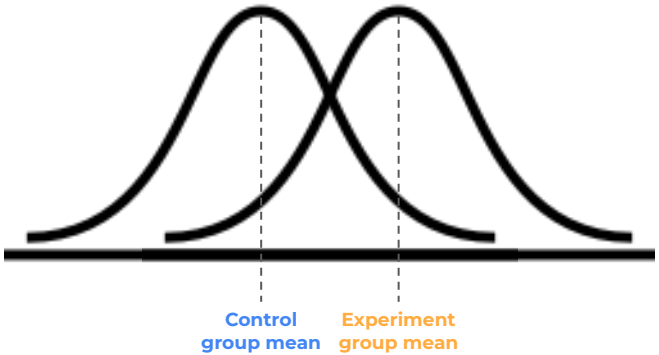
Introduction to Bayesian ML (what it means to be *Bayesian*)

-

Frequentist and Bayes comparison (How each one do the same things and why Bayes is better)

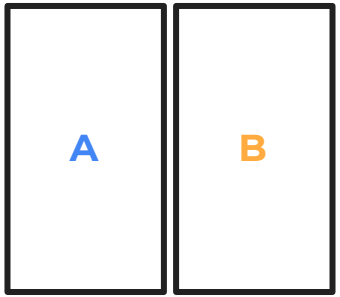
Examples

A new drug



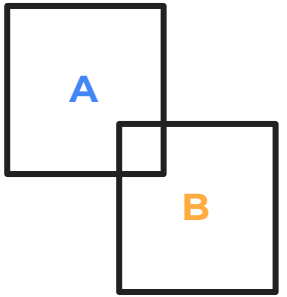
How do we know the experiment is actually performing better?

Website change



How to maximize the conversion rate?

Campaigns



Which performed better?

Comparison between two or more groups.
Each of these groups produce numbers.
What group has a higher/lower number (statistically speaking)?

These questions require calculations, estimations, modeling and statistics that we will cover in the course.
We will learn if a difference between groups is indeed *meaningful*

The Journey

Hour 2	Hour 3	Hour 4	Hour 5	Hour 6	Hour 7
Traditional AB Testing	**Bayes Rule**	**Bayesian Estimation**	**Density Functions**	**Bayesian AB Testing I**	**Bayesian AB Testing II**
Basic review and intuition on important concepts overlapping both approaches.	Statistics vs. Probability.	Distributions, randomness.	CDF.	Hands-on workshop.	Hands-on workshop.
Code: Confidence intervals, hypothesis testing, AB Test example	***Code: Probability***	***Code: Maximum Likelihood***	***Code: Percentiles***		

Github repository

1

From terminal, go to the folder you want the project to be in.

2

Create an environment for the project running:

```
python3 -m venv env
```

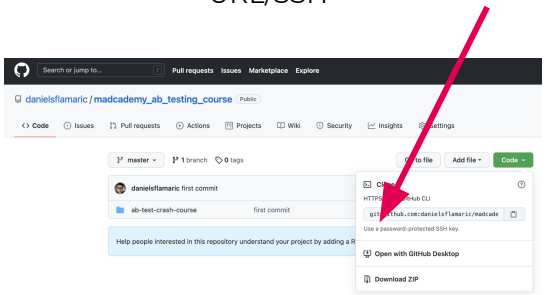
3

Activate the environment you just created:

```
source env/bin/activate
```

4

Go to the repo and copy its URL/SSH



5

Clone the repo into your project folder and open it.

```
git clone git@github.com:danielsflamaric/madcademy_ab_testing_course.git
```

```
madcademy_ab_testing_course % jupyter notebook
```