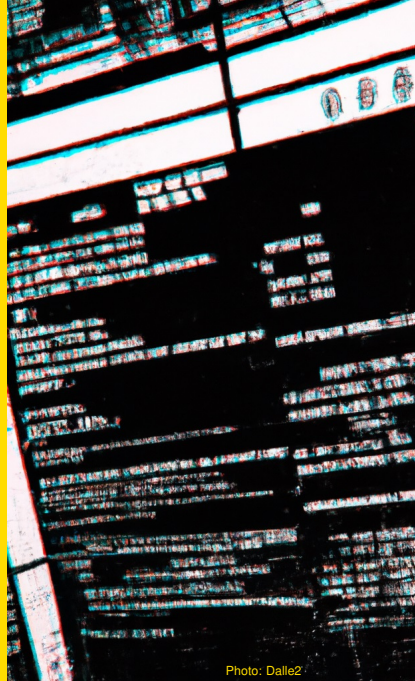# Data Mining

Inference 2 (Bayes)

Santiago Alonso-Díaz

Tecnólogico de Monterrey
EGADE, Business School

Data

Marketing :

- $Input_{costs}$, $Input_{marketing}$
- $Metrics_{CAC}$, $Metrics_{CLV}$
- $Output_{Costumers_{new}}$

CAC: consumer acquisition costs, CLV: consumer life value (in sales for the company)

Hypotheses $\neq$ Data

What generated $Costumers_{new}$?

- How about this model for the hypothesis that sales and marketing expenses work?

$$Costumers_{new} = \beta(Cost_{sales} + Cost_{marketing})$$

- β is not data (CAC could be a prior).
- β embodies our hypothesis that costs $\mapsto$ costumers.
- How can we estimate $\beta | data$?

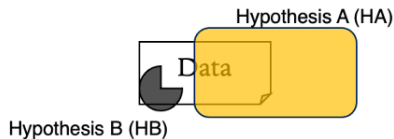# Table of contents

Bayesian Inference

Two inferential approaches

Hypothesis A (HA)

Data

Hypothesis B (HB)

Approach 1:
max. p(Data|H)
Given HB, the data covers most of HB. Pick HB

Approach 2:
max. p(H|Data)
Given the data, HA covers more data. Pick HA

Figure: Maximize data or hypotheses (or both).

What is Bayesian inference?

Estimate the probability of hypotheses given the data:

$$p(Hypotheses|Data)$$

In our example, we want $p(\beta|Costumers_{new}, Cost_{sales}, Cost_{marketing}$

$$Costumers_{new} = \beta(Cost_{sales} + Cost_{marketing})$$

In Bayesian inference we obtain a probability distribution of β.

But let's go to Python for the basics of Bayesian inference. The next sessions we would see how to obtain β of regressions (frequentist and bayesian).

Bayes and Marketing

Go to minute 16 to 26
Go to minute 3:55

References