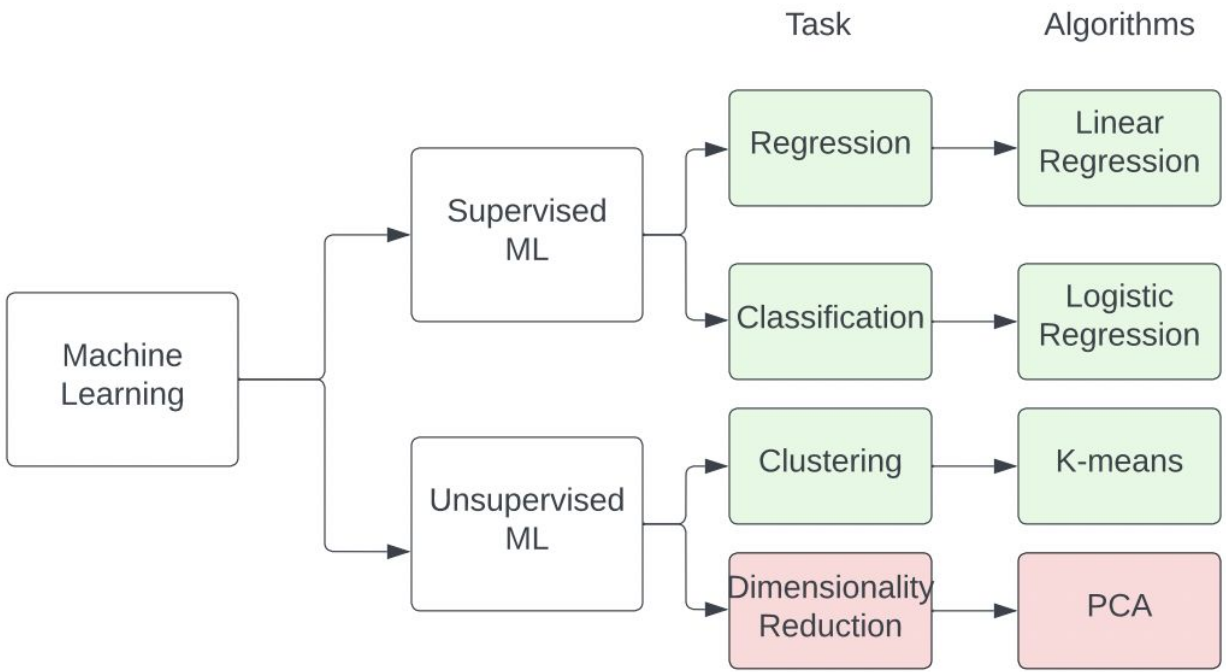
LECTURE 3: AI/ML BASIC CONCEPTS

Niek Mereu, ML engineer STRV

STRV

SCHEMATIC OVERVIEW

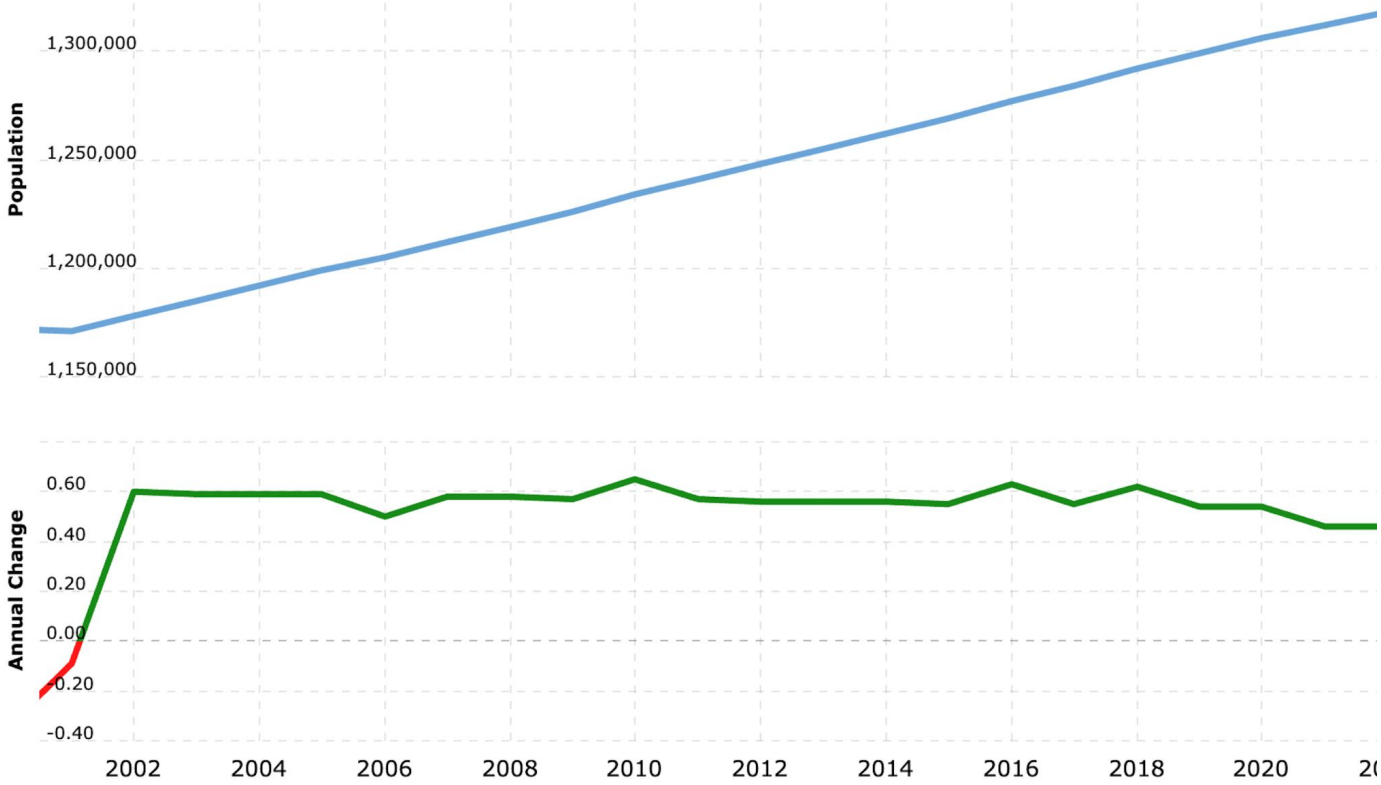


SUPERVISED ML

LINEAR REGRESSION & LOGISTIC REGRESSION

01

LINEAR RELATIONSHIP



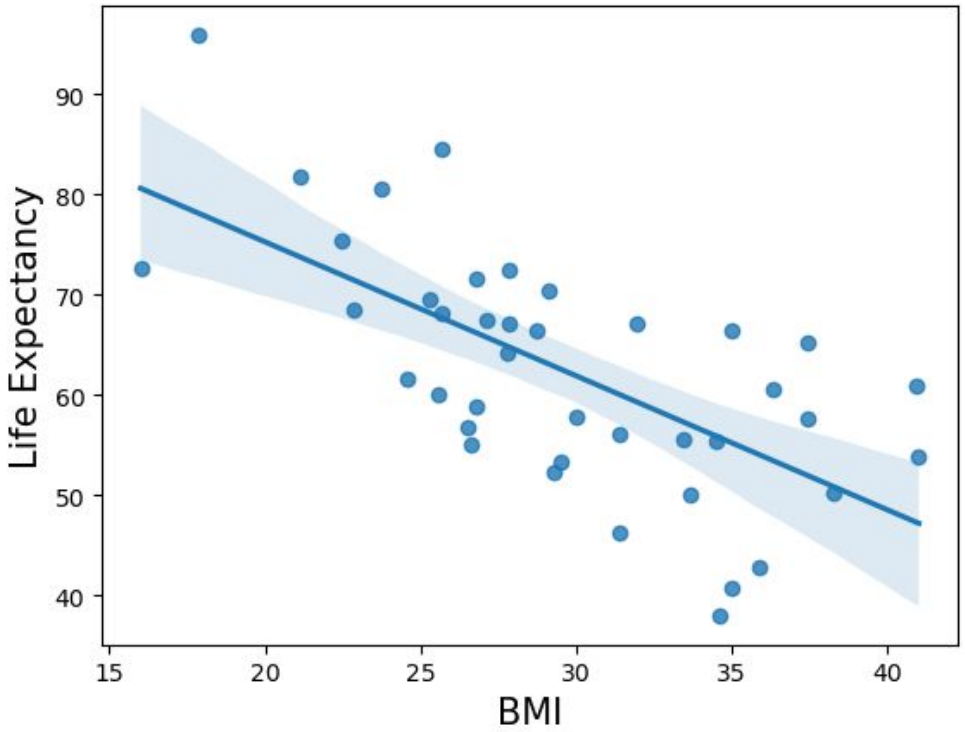
REGRESSION

- The goal is to predict Y, which is a real number.
- To predict Y, we use the collection of features X.
- Examples: height prediction, bank account forecasting…
- Algorithms: linear regression, random forest for regression, neural-network…

LINEAR REGRESSION

- Goal: predict target (y) given features (X).
- y is a numeric (real) number.
- Need to estimate a (intercept) and b (slope).

$$Y_{LE} = \alpha + \beta_{bmi} X_{bmi} + \epsilon$$



CLASSIFICATION

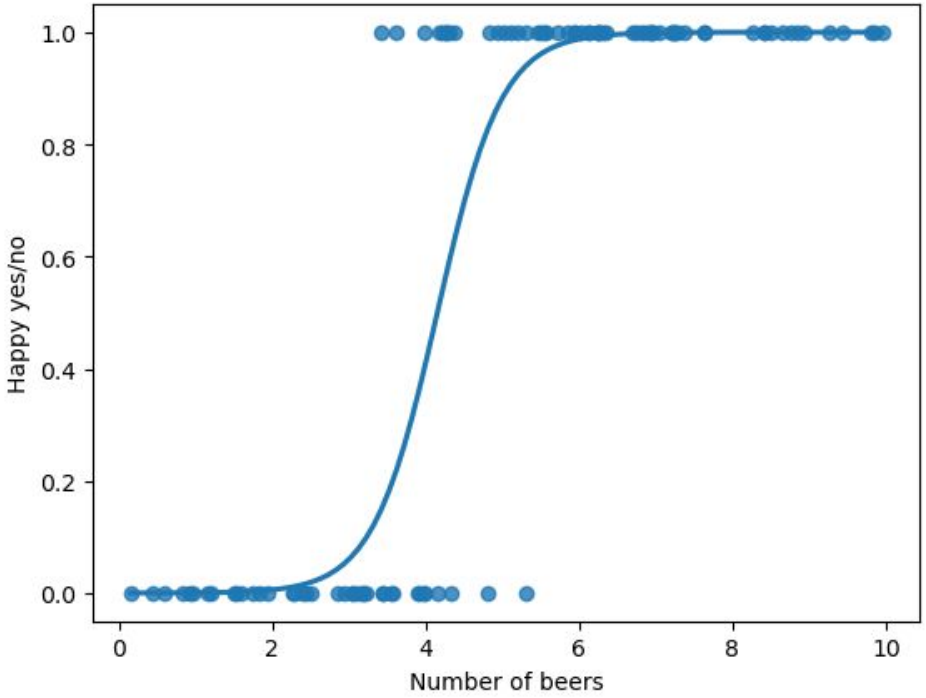
- We don’t want to predict a number(!)
- The response is now a label, representing a gender (male, female), animal (dog/cat)...
- Examples: e-mail spam filter, handwritten letters (a/b/c/…)
- Algorithms: Logistic regression, random-forest, neural network, naive-bayes…

LOGISTIC REGRESSION (CLASSIFICATION)

- Still y = a + bx!
- But linear function is
  Transformed to an S-shape.
- Output is a class, a label.
  The output is non-numeric.

$$f_{happy} = \alpha + \beta_{beer} X_{beer} + \epsilon$$

$$Y_{happy} = \sigma(f_{happy})$$

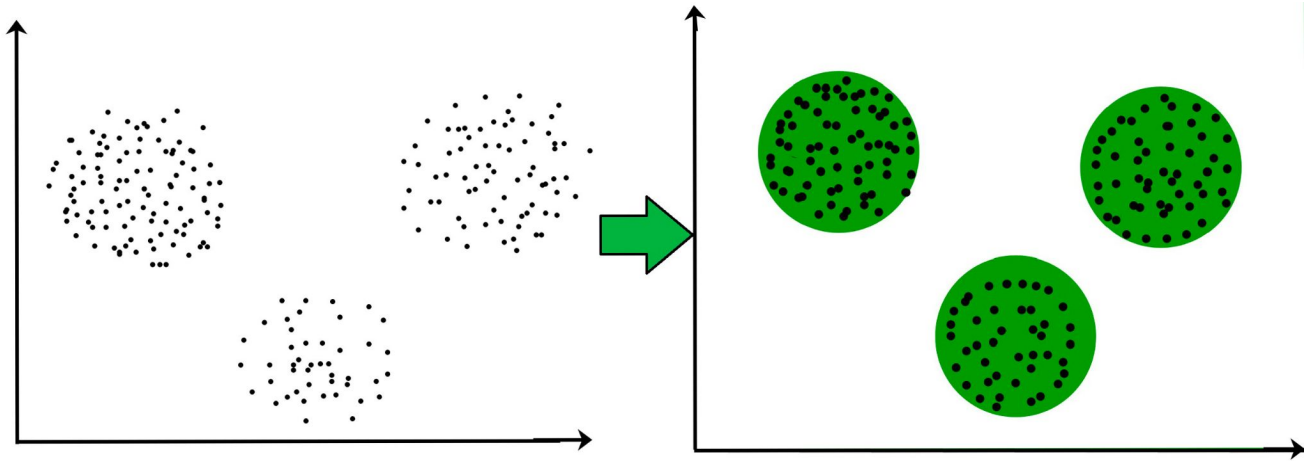


UNSUPERVISED ML
K-MEANS CLUSTERING

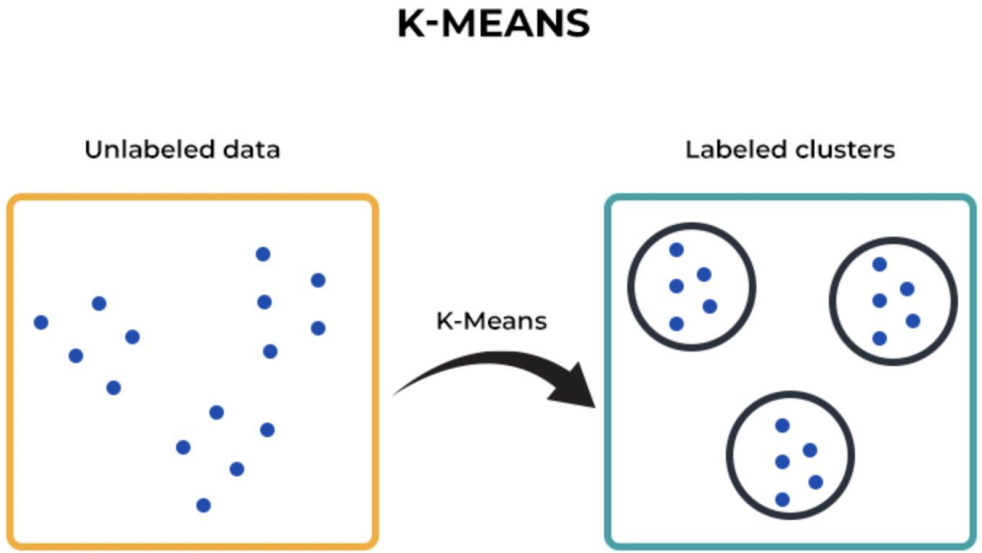
02

CLUSTERING

- Un-supervised so without Y, but with features X.
- Automatically discovering natural grouping in data
- Examples: finding customer groups, political party preference…



K-MEANS CLUSTERING



- Pick number of centers
- Put centers randomly in the space.
- Calculate closest points to centers
- Move center to the center closest points
- Repeat until centers are stable.
- Visualization: https://www.youtube.com/watch?v=5I3Ei69I40s