MACHINE LEARNING TOOLS

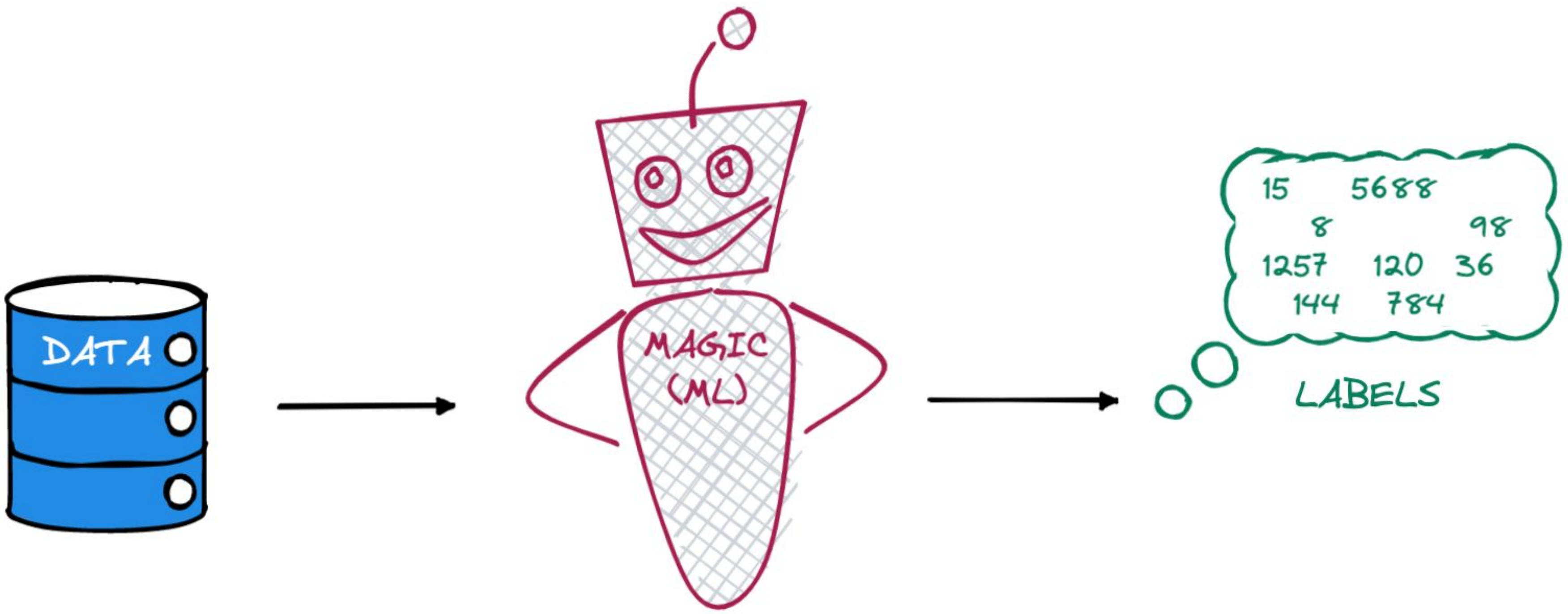
Central European University
2024

Photo by Markus Spiske on Unsplash

@divenyijanos

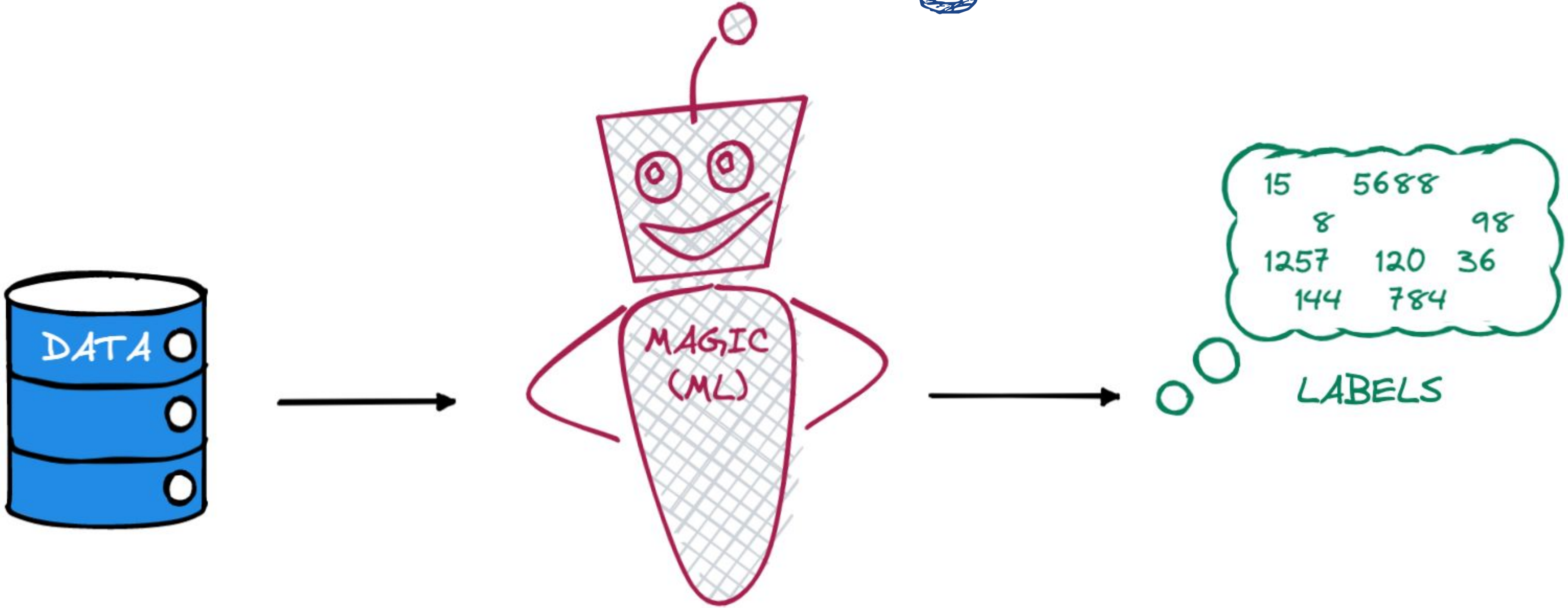

Google Cloud Platform
Introduction to ML & AI
Part 1 of Making Friends with Machine Learning
Episodes 000-028


MAGIC
(ML)



@divenyijanos

Machine Learning =



Thing Labeller

@divenyijanos

@divenyijanos


@divenyijanos

@divenyijanos

@divenyijanos

@divenyijanos

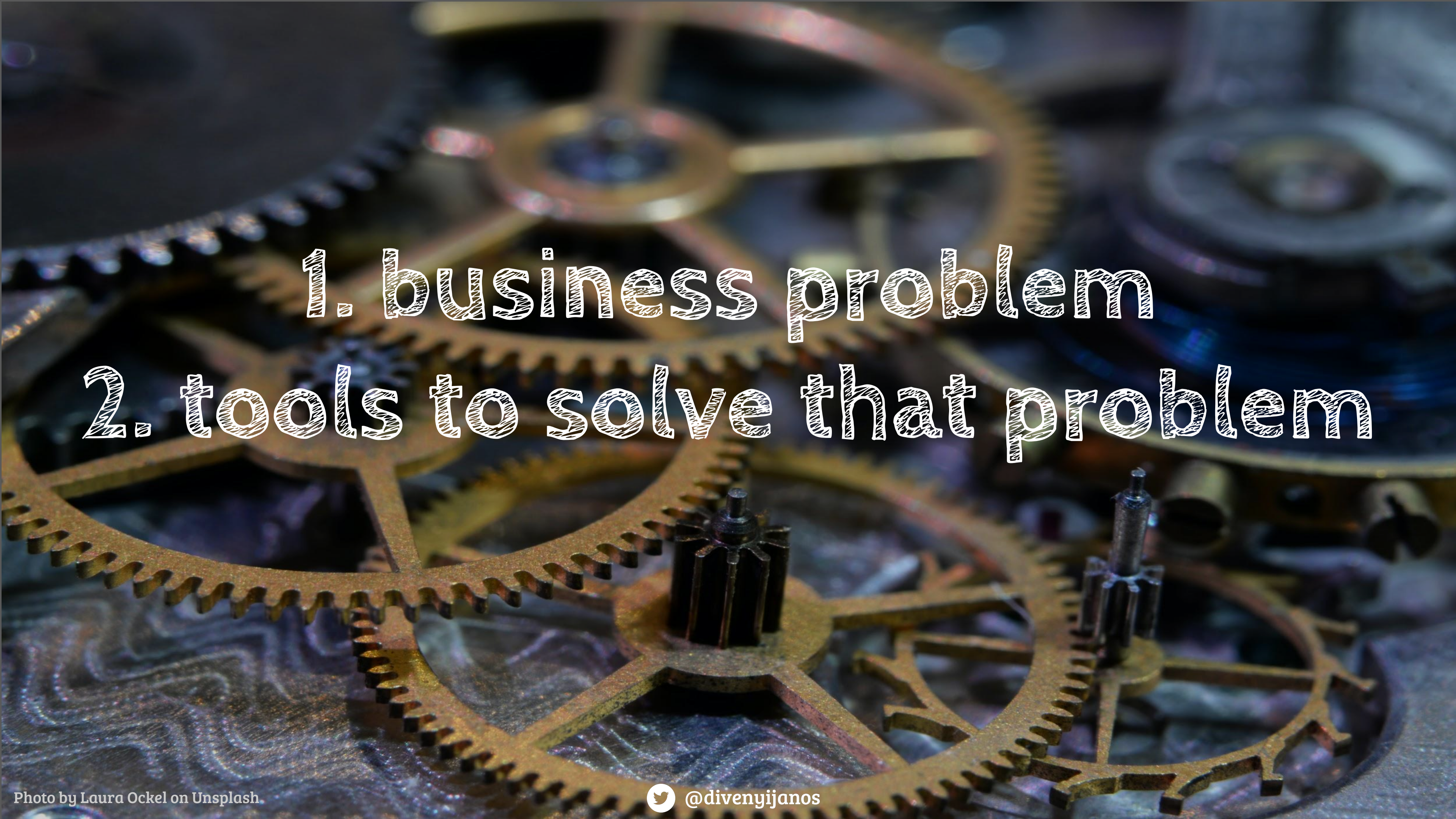
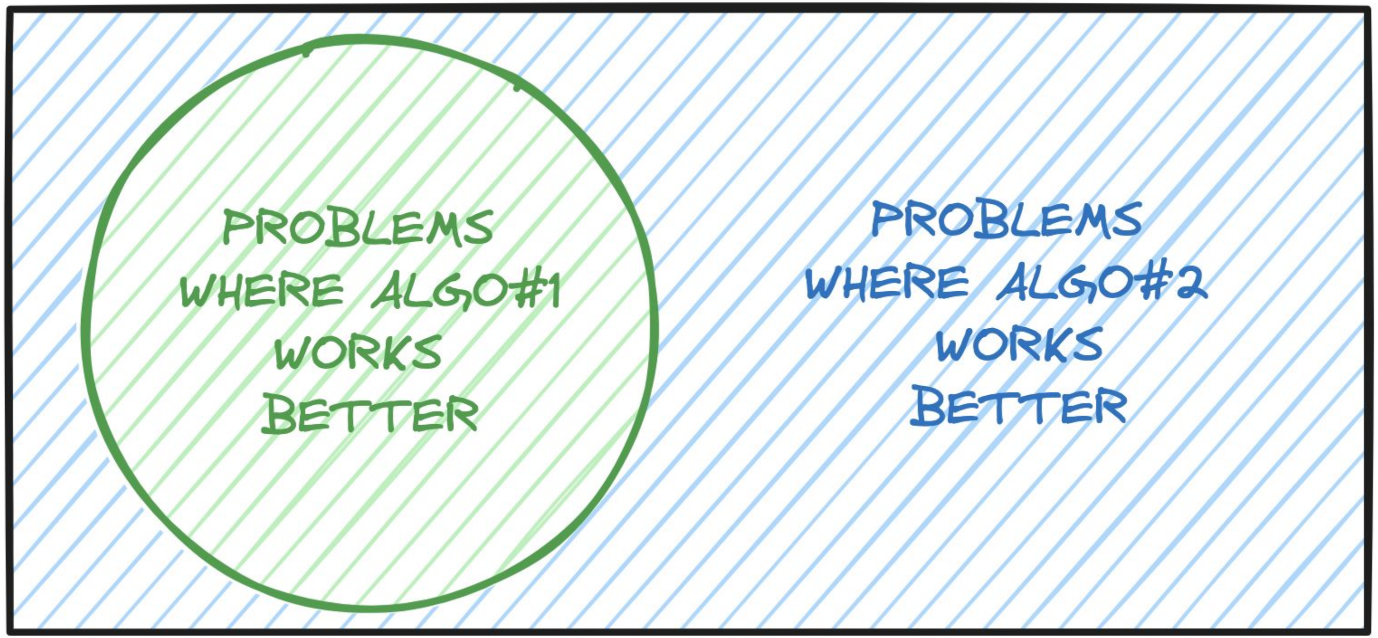


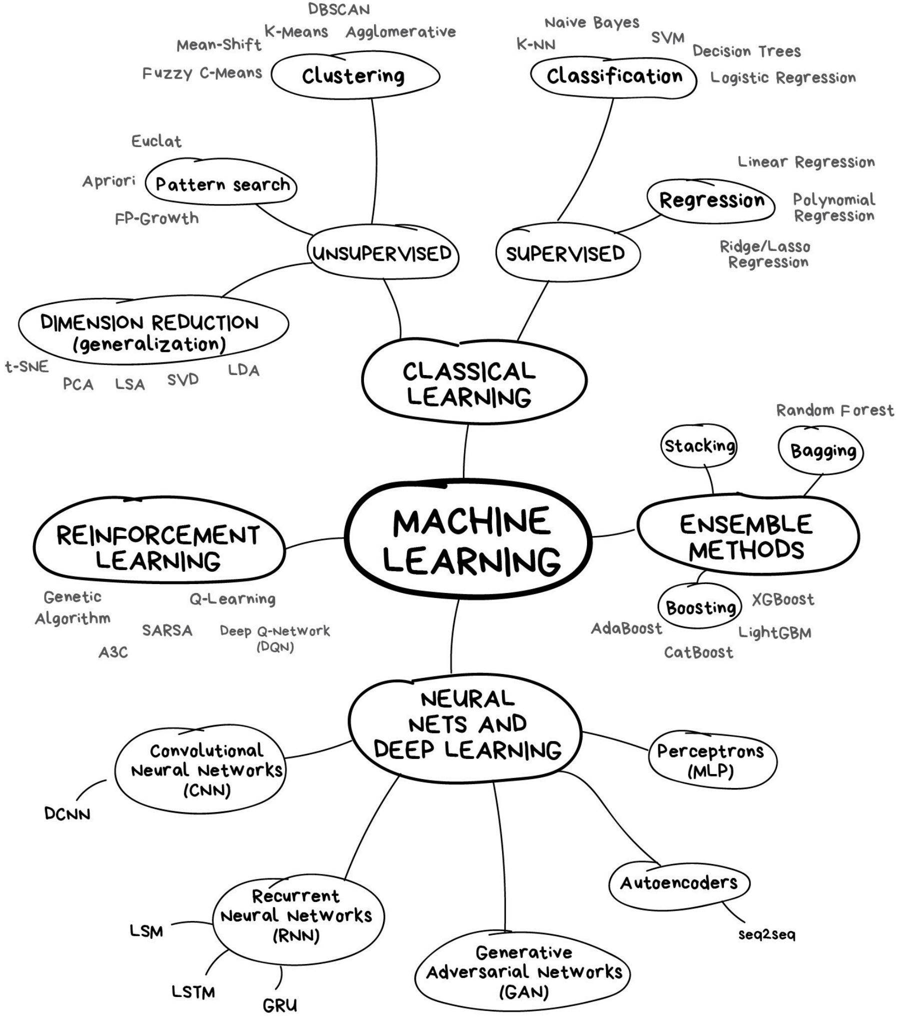
Photo by Laura Ockel on Unsplash

No Free Lunch theorem

ENTIRE SET OF PROBLEMS IN THE UNIVERSE

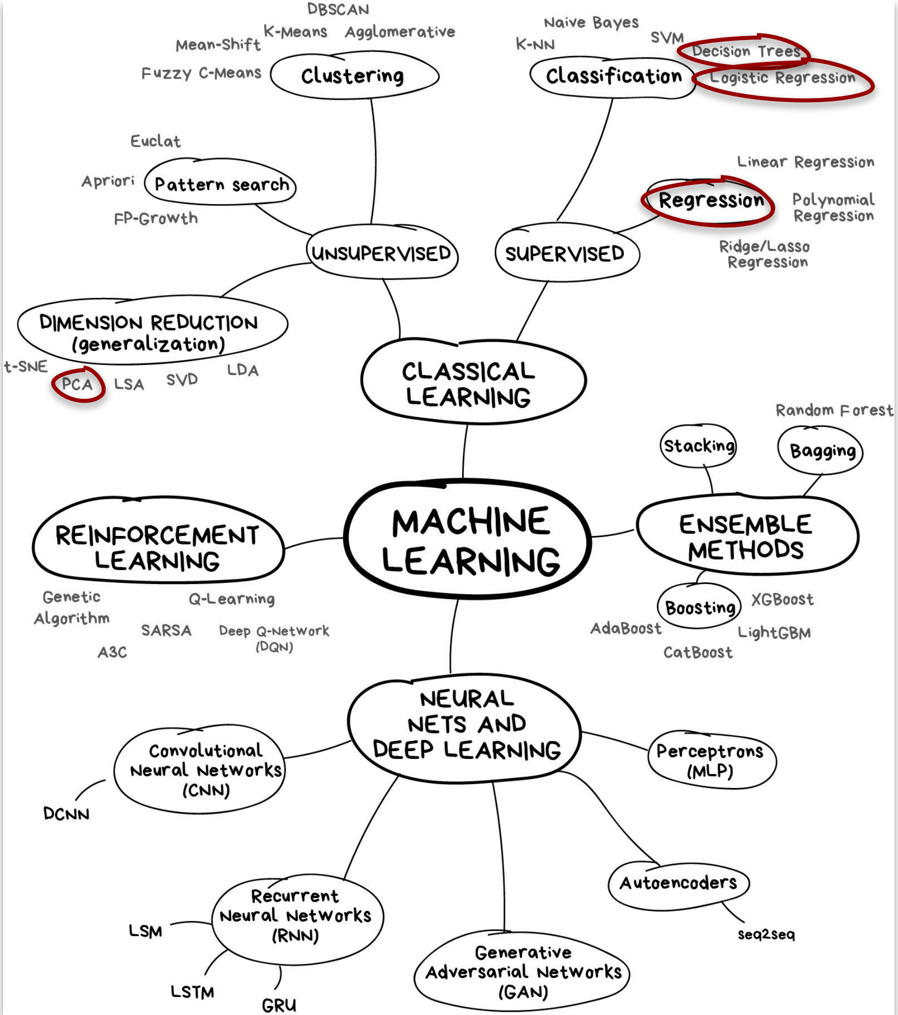


ASSUMING THERE ARE ONLY 2 ALGORITHMS



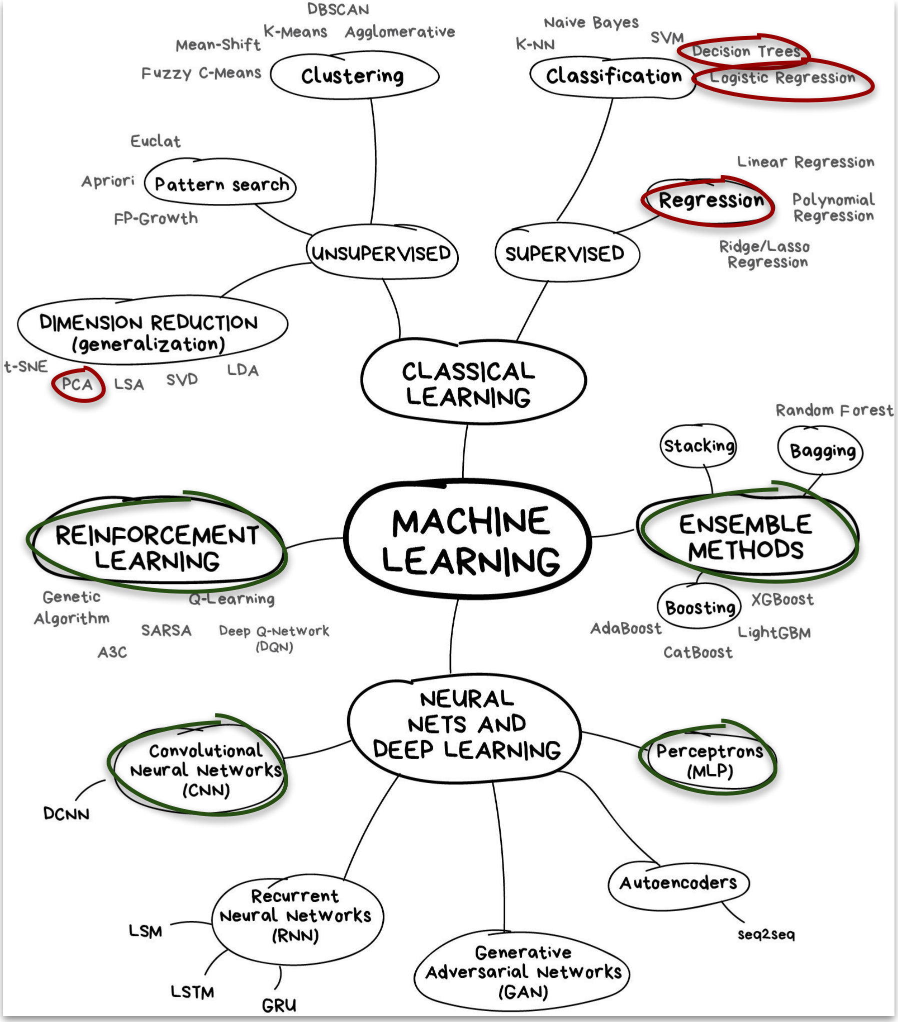
Drawing by Vastrik: https://vas3k.com/blog/machine_learning/

@divenyijanos



Drawing by Vastrik: https://vas3k.com/blog/machine_learning/

@divenyijanos



Drawing by Vastrik: https://vas3k.com/blog/machine_learning/

@divenyijanos

@divenyijanos

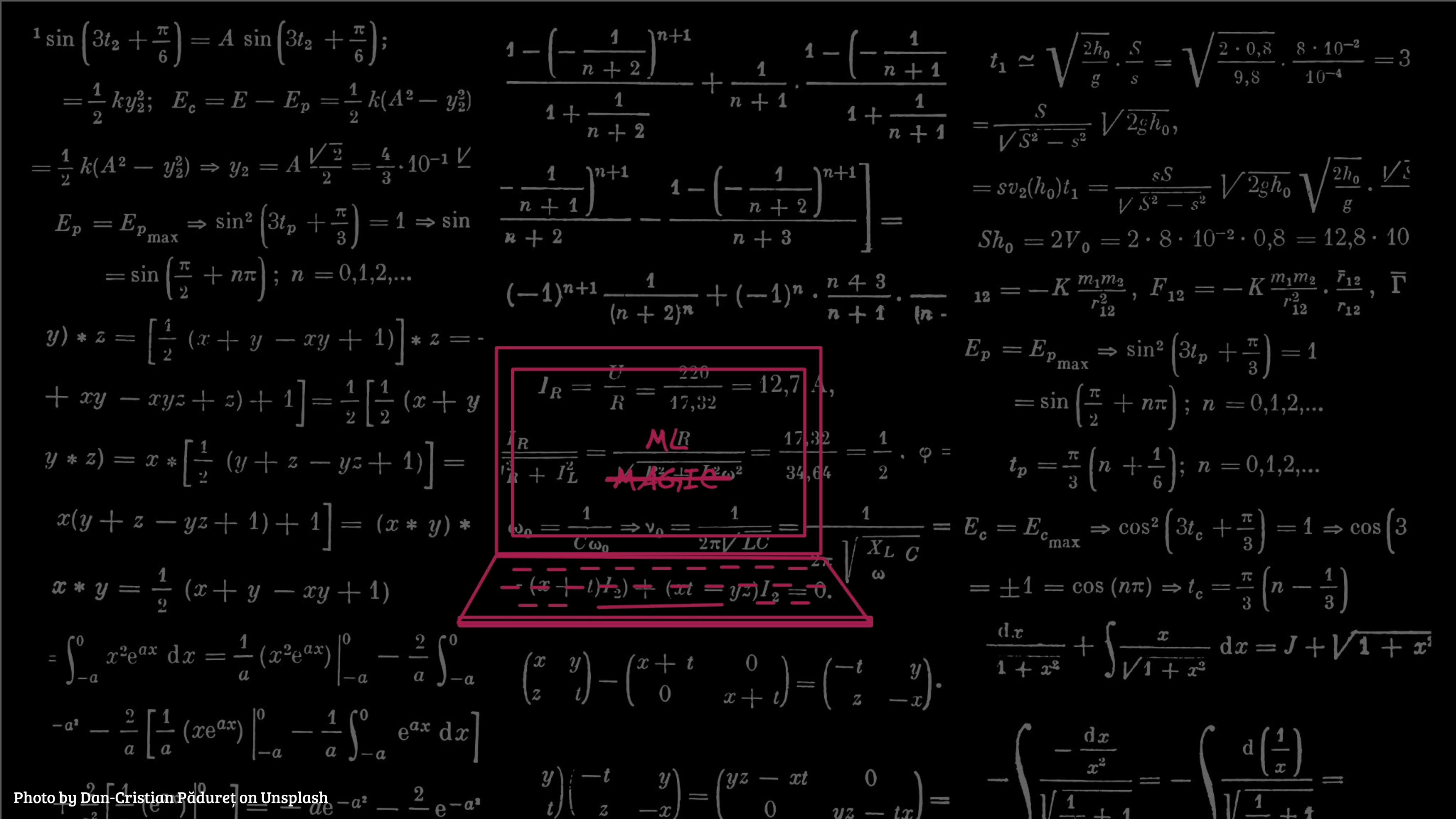


Photo by Dan-Cristian Pădureț on Unsplash

Photo by Dan-Cristian Pădureț on Unsplash

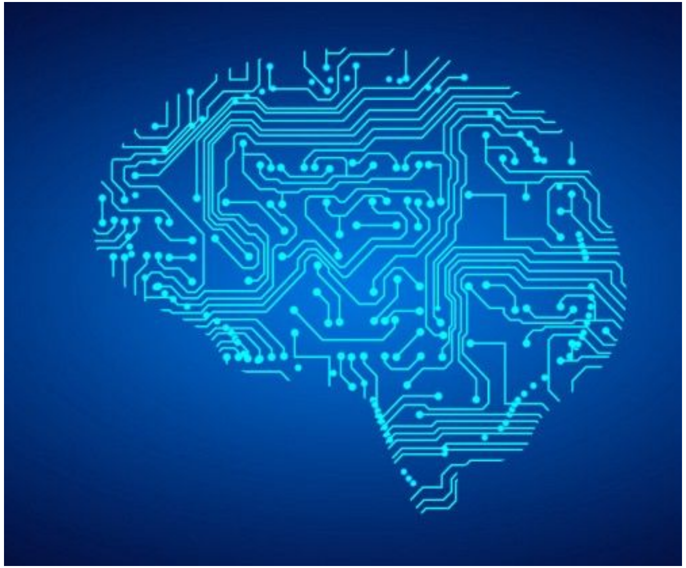
GIF made from video by @kareem_carr


DATA
ML
~~MAGIC~~
15
5688
8
98
1257
120
36
144
784
LABELS

Loss function predicted vs real (e.g. MSE)

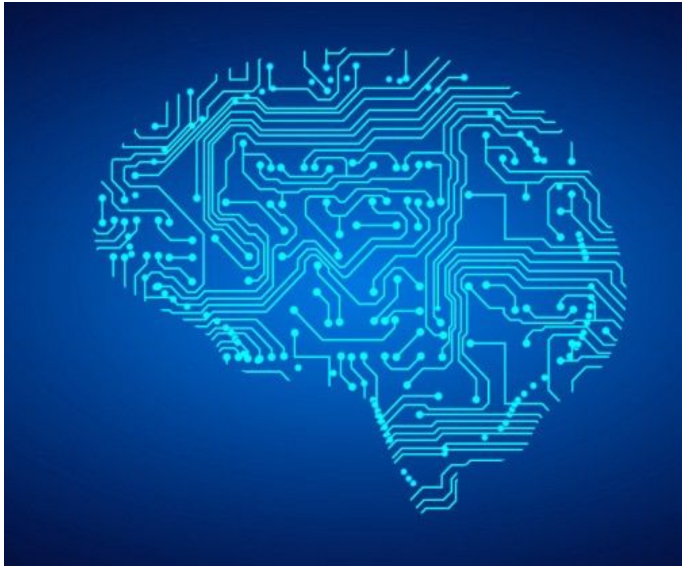
Loss function predicted vs real depending on params (e.g. MSE)
Optimization algorithm tweak params to minimize loss (e.g. derivation)

Loss function predicted vs real depending on params (e.g. MSE)
Optimization algorithm tweak params to minimize loss (e.g. derivation)



Memorization vs Generalization

Loss function predicted vs real depending on params (e.g. MSE)
Optimization algorithm tweak params to minimize loss (e.g. derivation)



~~Memorization~~ vs Generalization

Loss function predicted vs real depending on params (e.g. MSE)
Optimization algorithm tweak params to minimize loss (e.g. derivation)

~~Memorization~~ vs Generalization

Loss function predicted vs real depending on params (e.g. MSE)
Optimization algorithm tweak params to minimize loss (e.g. derivation)
+regularization prefer simpler models → optimize for Loss + Penalty

Loss function predicted vs real depending on params (e.g. MSE)
Optimization algorithm tweak params to minimize loss (e.g. derivation)
+regularization prefer simpler models → optimize for Loss + Penalty
+hyperparameters control penalty + algo details

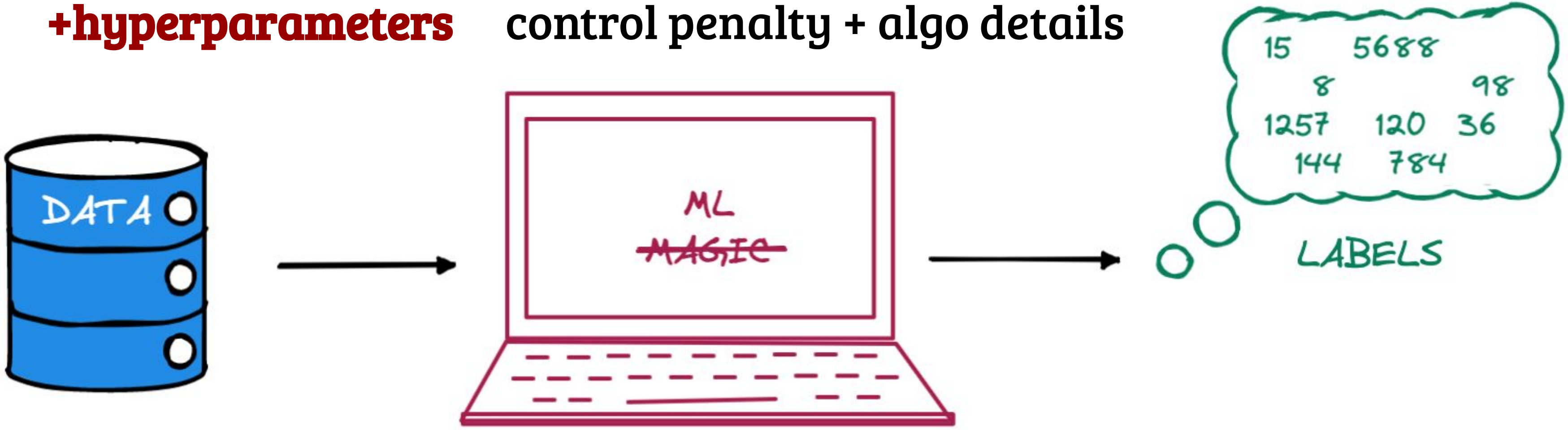


Photo by Laura Ockel on Unsplash

@divenyijanos

Miscellaneous

- Materials on Moodle: ~~mandatory~~ recommended reading & videos
- Discussion in Slack: help each other!
- Grading
  - 30% Quizzes (3x10): 7 test questions on Moodle 13:30-13:45
    - points = #correct answers + 5 (no answer: 0)
  - 40% Assignments (2x20)
    - practical tasks that require coding + explanation
    - `.ipynb` and rendered `.pdf` expected (2 separate files!)
    - deadlines: Friday 8PM of 22 March and 5 April
  - 30% Take home exam
    - Kaggle competition
    - submission + `.ipynb` and rendered `.pdf` expected (2 separate files!)
    - deadline: Friday 8PM of 19 April

Demand Forecast for Bike Share

👉 in Washington DC...

@divenyijanos

Recommended Materials

Video:

- Cassie Kozyrkov (Google): Introduction to ML and AI (recommended)
- Josh Starmer (StatQuest): Random Forests Part 1
- Josh Starmer (StatQuest): AdaBoost, Clearly Explained
- Josh Starmer (StatQuest): Gradient Boost Part 1

Text:

- r2d3: Visual explanation of how tree model works
- ISLR 8.2: Bagging, Random Forests, Boosting, and Bayesian Additive Regression Trees
- ESL 10: Boosting and Additive Trees
- ESL 15: Random Forests