

INTRODUÇÃO A MACHINE LEARNING

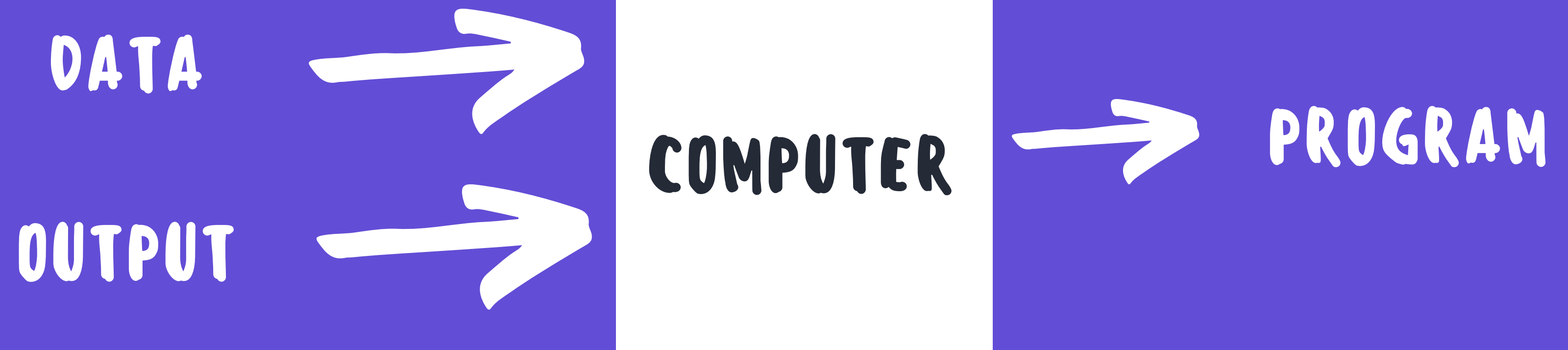
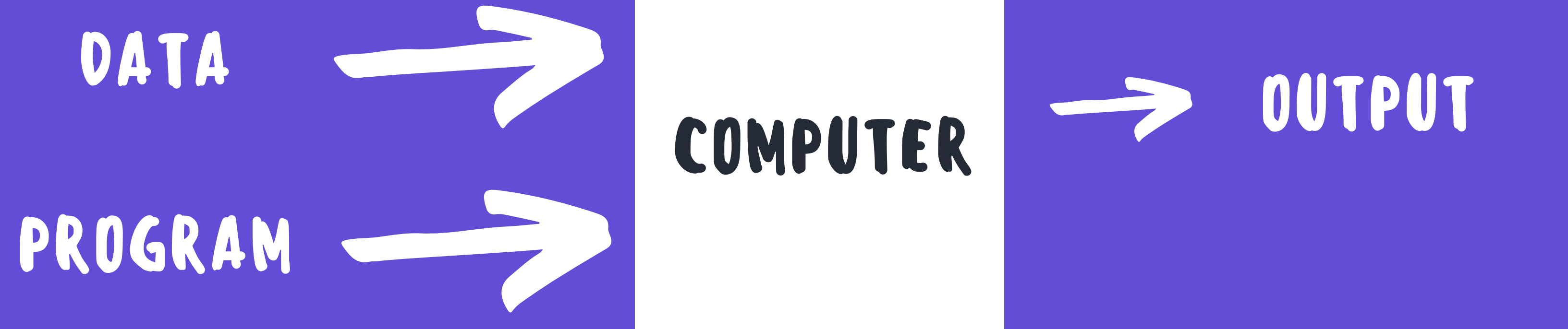
WHAT IS MACHINE LEARNING?

"LEARNING IS ANY PROCESS BY WHICH A SYSTEM IMPROVES PERFORMANCE FROM EXPERIENCE" - HERBERT SIMON

DEFINITION (1998 - TOM MITCHELL):

MACHINE LEARNING IS THE STUDY OF ALGORITHMS THAT

- IMPROVE THEIR PERFORMANCE P
- AT SOME TASK T
- WITH EXPERIENCE E



WHEN DO WE USE MACHINE LEARNING?

ML IS USED WHEN:

- HUMAN EXPERTISE DOES NOT EXIST
- HUMANS CAN'T EXPLAIN THEIR EXPERTISE
- MODELS MUST BE CUSTOMIZED
- MODELS ARE BASED ON HUGE AMOUNTS OF DATA

EX.: NAVIGATING ON MARS, SPEECH RECOGNITION, PERSONALIZED MEDICINE, GENOMICS

SOME MORE EXAMPLES OF TASKS THAT ARE BEST SOLVED BY USING A LEARNING ALGORITHM

- RECOGNIZING PATTERNS:
 - FACIAL IDENTITIES OF FACIAL EXPRESSIONS
 - HANDWRITTEN OR SPOKEN WORDS
 - MEDICAL IMAGES
- GENERATING PATTERNS:
 - GENERATING IMAGES OR MOTION SEQUENCES
- RECOGNIZING ANOMALIES:
 - UNUSUAL CREDIT CARD TRANSACTIONS
 - UNUSUAL PATTERNS OF SENSOR READINGS IN A NUCLEAR POWER PLANT

DEFINING THE LEARNING TASK

IMPROVE ON TASK T, WITH RESPECT TO PERFORMANCE METRIC P,
BASED ON EXPERIENCE E

EX.:

T: CATEGORIZE EMAIL MESSAGES AS SPAM OR LEGITIMATE

P: PERCENTAGE OF EMAIL MESSAGES CORRECTLY CLASSIFIED

E: DATABASE OF EMAILS, SOME WITH HUMAN-GIVEN LABELS

EX2.:

T: RECOGNIZING HAND-WRITTEN WORDS

P: PERCENTAGE OF WORDS CORRECTLY CLASSIFIED

E: DATABASE OF HUMAN-LABELED IMAGES OF HANDWRITTEN WORDS

O QUE VAMOS VER:

- PYTHON
- DATA PREPROCESSING

- REGRESSION:

- SIMPLE LINEAR REGRESSION
- MULTIPLE LINEAR REGRESSION
- POLYNOMIAL REGRESSION
- SUPPORT VECTOR REGRESSION (SVR)
- DECISION TREE REGRESSION
- RANDOM FOREST REGRESSION
- EVALUATING REGRESSION MODELS PERFORMANCE

- **CLASSIFICATION:**

- **LOGISTIC REGRESSION**
- **K-NEAREST NEIGHBORS (K-NN)**
- **SUPPORT VECTOR MACHINE (SVM)**
- **KERNEL SVM**
- **NAIVE BAYES**
- **DECISION TREE CLASSIFICATION**
- **RANDOM FOREST CLASSIFICATION**
- **EVALUATING CLASSIFICATION MODELS PERFORMANCE**

- CLUSTERING
 - HIERARCHICAL CLUSTERING
 - APRIORI
 - ECLAT
 - UPPER CONFIDENCE BOUND (UCB)
 - TROMPSON SAMPLING
-
- NATURAL LANGUAGE PROCESSING
 - DEEP LEARNING
 - ARTIFICIAL NEURAL NETWORKS
 - CONVOLUTIONAL NEURAL NETWORKS
 - PRINCIPAL COMPONENT ANALYSIS (PCA)
 - LINEAR DISCRIMINANT ANALYSIS (LDA)