

# Introduction to machine learning

Alexis Zubiolo

Adcash Bulgaria

September 15, 2016

**Disclaimer:** This presentation is a promotion of the *Introduction to machine learning* course I will be giving at IT STEP.

# Presentation outline

- ▶ Introduction: What is machine learning?
- ▶ A few practical examples
  - ▶ classification
  - ▶ regression
- ▶ Goals and presentation of the course
- ▶ Questions and answers

# What is machine learning?

Let's start with a simple example...



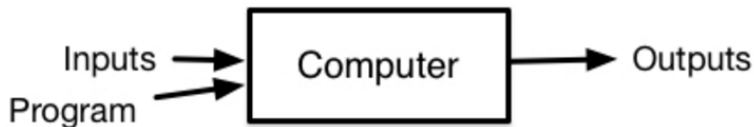
How to filter spam emails **automatically**?

# Machine learning paradigm

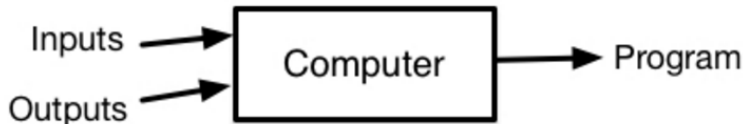
Goal: Build algorithms that can

- ▶ **learn** from data
- ▶ **make predictions** on data

## Traditional Programming



## Machine Learning



# Main components of machine learning

## Mathematics

- ▶ Linear algebra
- ▶ Calculus
- ▶ Numerical optimization

## Statistics, probability theory

## Computer science

## Example 1: Regression

Regression = output is a **continuous** numerical value

Example: **Estimate the price** of an apartment

- ▶ input: **information** about the apartment
- ▶ output: **price**

## Example 1: Regression

Regression = output is a **continuous** numerical value

Example: **Estimate the price** of an apartment

- ▶ input: **information** about the apartment
- ▶ output: **price**

living area (m <sup>2</sup> )	price (1000's euros)
50	30
76	48
26	12
102	90



## Example 1: Regression

Regression = output is a **continuous** numerical value

Example: **Estimate the price** of an apartment

- ▶ input: **information** about the apartment
- ▶ output: **price**

living area (m <sup>2</sup> )	price (1000's euros)
50	30
76	48
26	12
102	90
61	?

Linear model:  $\text{price} = \mathbf{a} \times \text{area} + \mathbf{b}$

Problem: optimal values for **a** and **b**?

# Regression

More data for a richer model:

living area (m <sup>2</sup> )	# bedrooms	price (1000's euros)
50	<b>1</b>	30
76	<b>2</b>	48
26	<b>1</b>	12
102	<b>3</b>	90
61	<b>2</b>	?

Linear model:  $\text{price} = \mathbf{a} \times \text{area} + \mathbf{b} \times \# \text{ bedrooms} + \mathbf{c}$

Problem: optimal values for **a**, **b** and **c**?

## Example 2: Classification

Classification = output is a **label**

Examples:

## Example 2: Classification

Classification = output is a **label**

Examples:

- ▶ Spam filtering
  - ▶ input: email (text, subject, address, ...)
  - ▶ output: **spam** or **not spam**

## Example 2: Classification

Classification = output is a **label**

Examples:

- ▶ Spam filtering
  - ▶ input: email (text, subject, address, ...)
  - ▶ output: **spam** or **not spam**
- ▶ Object recognition in images or videos
  - ▶ input: image or video
  - ▶ (example) output: **face** or **not a face**

## Example 2: Classification

Classification = output is a **label**

Examples:

- ▶ Spam filtering
  - ▶ input: email (text, subject, address, ...)
  - ▶ output: **spam** or **not spam**
- ▶ Object recognition in images or videos
  - ▶ input: image or video
  - ▶ (example) output: **face** or **not a face**
- ▶ Image classification/description
  - ▶ input: image
  - ▶ output: image **description** or **label** (apple, car, ...)

# Automated image description generation



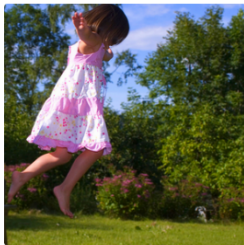
"man in black shirt is playing guitar."



"construction worker in orange safety vest is working on road."



"two young girls are playing with lego toy."



"girl in pink dress is jumping in air."



"black and white dog jumps over bar."



"young girl in pink shirt is swinging on swing."

## Automated image description generation





## Automated image description generation



A man holding a red apple in his mouth

## Other topics

Machine learning is a wide and growing field. It also includes:

## Other topics

Machine learning is a wide and growing field. It also includes:

- ▶ Clustering (no predefined label/output)

## Other topics

Machine learning is a wide and growing field. It also includes:

- ▶ Clustering (no predefined label/output)
- ▶ Dimensionality reduction

## Other topics

Machine learning is a wide and growing field. It also includes:

- ▶ Clustering (no predefined label/output)
- ▶ Dimensionality reduction
- ▶ ...

# The course

Important note: This is an **introduction** course.

# The course

Important note: This is an **introduction** course.

Goals:

- ▶ Introduce **main concepts** of machine learning
- ▶ **Implement** these concepts in Python (Scikit-learn)

# The course

Important note: This is an **introduction** course.

Goals:

- ▶ Introduce **main concepts** of machine learning
- ▶ **Implement** these concepts in Python (Scikit-learn)

Practical information:

- ▶ ~ **10 sessions**, 1 per week (early October – mid December)
- ▶ ~ **90min** sessions
- ▶ Thursdays at ~ 6:30pm
- ▶ Alternating between **lectures and lab sessions**



Thank you! Questions?