Machine Learning

Motivation behind Machine Learning

Sometimes we encounter problems for which it's really hard to write a computer program to solve. For example, let's say we wanted to program a computer to recognize hand-written digits:

You could imagine trying to devise a set of rules to distinguish each individual digit. Zeros, for instance, are basically one closed loop. But what if the person didn't perfectly close the loop. Or what if the right top of the loop closes below where the left top of the loop starts?

A zero that's difficult to distinguish from a six

In this case, we have difficulty differentiating zeros from sixes.

What is ML

In Very Simple Word:

Machine learning is a core sub-area of artificial intelligence; it enables computers to get into a mode of self-**learning** without being explicitly programmed. When exposed to new data, these computer programs are enabled to learn, grow, change, and develop by themselves

- Automating automation

- Getting computers to program themselves

- Writing software is the bottleneck

- Let the data do the work instead!

Traditional Programming

Data $\rightarrow$

Program $\rightarrow$

Computer

$\rightarrow$ Output

Machine Learning

Data $\rightarrow$

Output $\rightarrow$
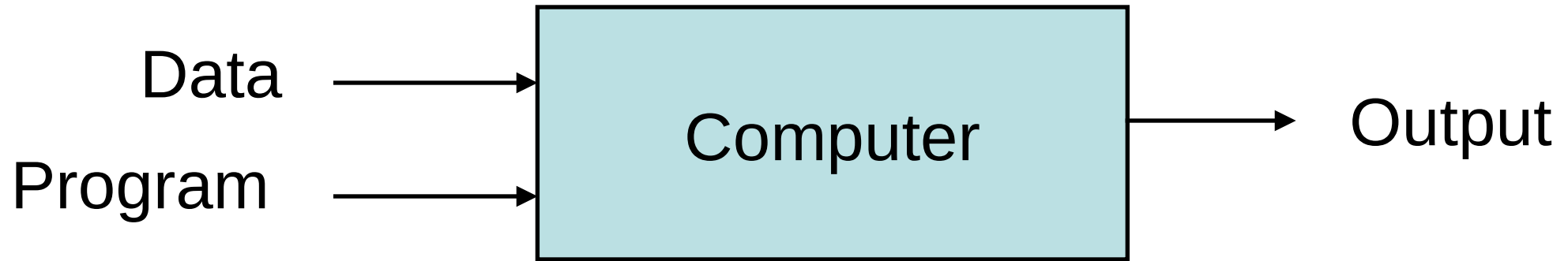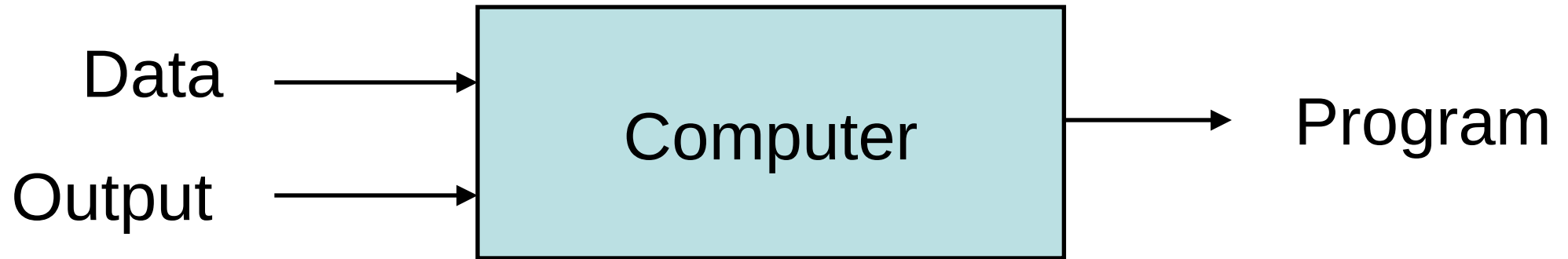
Computer

$\rightarrow$ Program

Magic?

No, more like gardening

- **Seeds** = Algorithms
- **Nutrients** = Data
- **Gardener** = You
- **Plants** = Programs

Why "Learn"?

- The key idea is to let a computer learn by example data or past experience instead of programming it with specific rules.
- There is no need to "learn" to calculate payroll
- Machine learning is a way of teaching computers to recognize patterns, and it's particularly useful in making sense of large amounts of data.
- Learning is used when:
  - Human expertise does not exist (navigating on Mars),
  - Humans are unable to explain their expertise (speech recognition)
  - Solution changes in time (routing on a computer network)
  - Solution needs to be adapted to particular cases (user biometrics)

A Few Quotes

"A breakthrough in machine learning would be worth ten Microsofts"

(Bill Gates, Chairman, Microsoft)

"Machine learning is the next Internet"

(Tony Tether, Director, DARPA)

Machine learning is the hot new thing"

(John Hennessy, President, Stanford)

"Web rankings today are mostly a matter of machine learning"

(Prabhakar Raghavan, Dir. Research, Yahoo)

"Machine learning is going to result in a real revolution"

(Greg Papadopoulos, CTO, Sun)

"Machine learning is today's discontinuity"

(Jerry Yang, CEO, Yahoo)

What We Talk About When We Talk About"Learning"

- Learning general models from a data of particular examples
- Data is cheap and abundant (data warehouses, data marts); knowledge is expensive and scarce.
- Example in retail: Customer transactions to consumer behavior:

    People who bought "Da Vinci Code" also bought "The Five People You Meet in Heaven" (www.amazon.com)

- Build a model that is *a good and useful approximation* to the data.

Sample Applications

- Web search

- Medical outcomes analysis

- Speech recognition

- Computer vision

- Robot control

- Computational biology

- Finance

- E-commerce

- Space exploration

- Robotics

- Information extraction: text to records

- Social networks

- Natural language processing

Types of ML

The field itself: ML is a field of study which harnesses principles of computer science and statistics to create statistical models. These models are generally used to do two things:

- **Prediction:**

    make predictions about the future based on data about the past

- **Inference:**

    discover patterns in data

ML in a Nutshell

- Tens of thousands of machine learning algorithms
- Hundreds new every year -predictive and classification type
- Every machine learning algorithm has three components:

    - Representation (landscape of possible models)

    - Evaluation (how do you prefer one model over the other)

    - Optimization (finetuning to find best performing model)

ML v/s AI

- **Artificial Intelligence (AI)** and **Machine Learning (ML)** are two very hot buzzwords right now, and often seem to be used interchangeably.

- They are not quite the same thing, but the perception is that they are

Artificial Intelligence is the broader concept of machines being able to carry out tasks in a way that we would consider "smart".

Machine Learning is a current application of AI based around the idea that we should really just be able to give machines access to data and let them learn for themselves.

ML: Decision tree learning

- **Decision tree learning** uses a decision tree (as a predictive model) to go from observations about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves)

- It is one of the predictive modeling approaches used in statistics, data mining and machine learning

- Tree models where the target variable can take a discrete set of values are called **classification trees**

- Decision trees where the target variable can take continuous values (typically real numbers) are called **regression trees**.
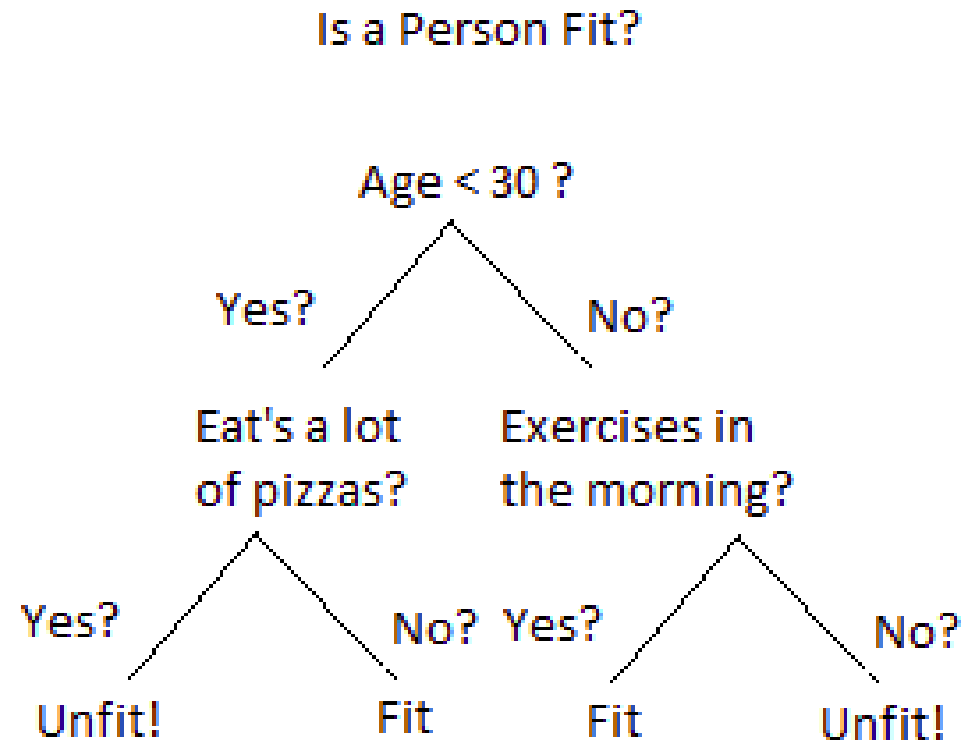
Example 1

- A tree showing survival of passengers on the _Titanic_. The figures under the leaves show the probability of survival and the percentage of observations in the leaf.
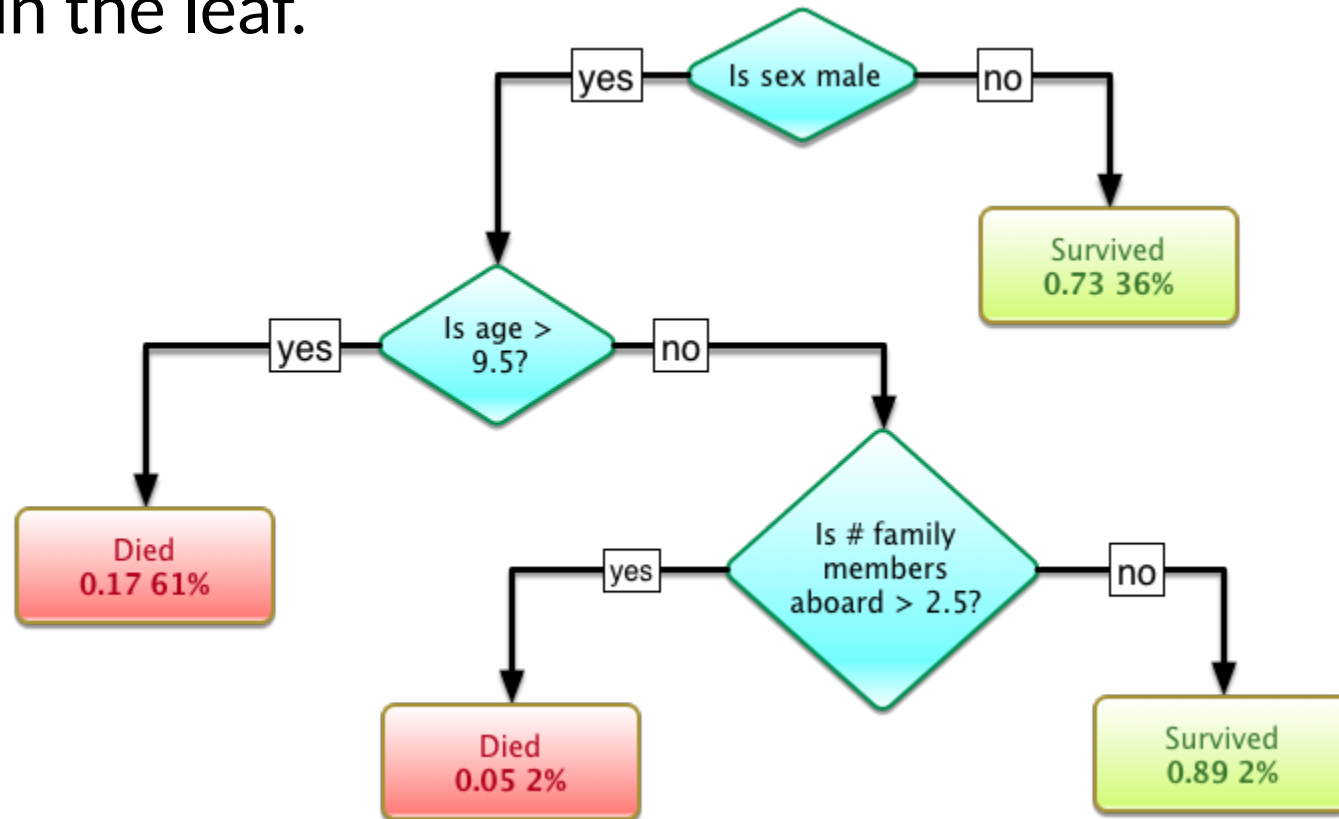
Example 2

- Let's say you want to predict whether a person is fit given their information like age, eating habit, and physical activity, etc.

Is a Person Fit?

Age < 30 ?

Yes?   No?

Eat's a lot   Exercises in
of pizzas?   the morning?

Yes?   No? Yes?   No?

Unfit!   Fit   Fit   Unfit!

What is the best programming language for machine learning?

Python.

Python if a popular **scientific** language and a rising star for machine learning. I'd be surprised if it can take the data analysis mantle from R, but matrix handling in **NumPy** may challenge MATLAB and communication tools like Ipython (interactive computing) are very attractive and a step into the future of reproducibility.

Types of machine learning Algorithms

- Supervised learning
- Unsupervised Learning
- Semi-supervised Learning
- Reinforcement Learning

Supervised Learning

- The Figure 1 shows an example of a supervised learning process used to produce a model capable of recognize ducks in images.
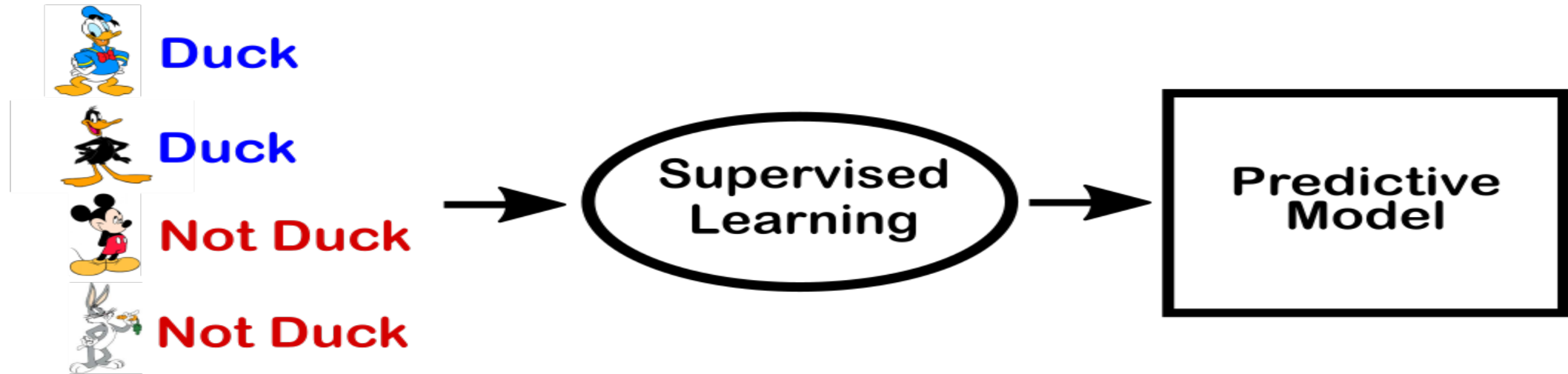
- Once trained, the resulting predictive model can be deployed to a production environment (a mobile app, for example) in order to recognize new pictures, as shown in Figure 2.

Unsupervised Machine Learning

- Unsupervised learning is where you only have input data (X) and no corresponding output variables.

- The goal for unsupervised learning is to model the underlying structure or distribution in the data in order to learn more about the data.

- These are called unsupervised learning because unlike supervised learning above there is no correct answers and there is no teacher. Algorithms are left to their own devises to discover and present the interesting structure in the data

Unsupervised Machine Learning

Unsupervised learning is a type of machine learning algorithm used to draw inferences from datasets consisting of input data without labeled responses.
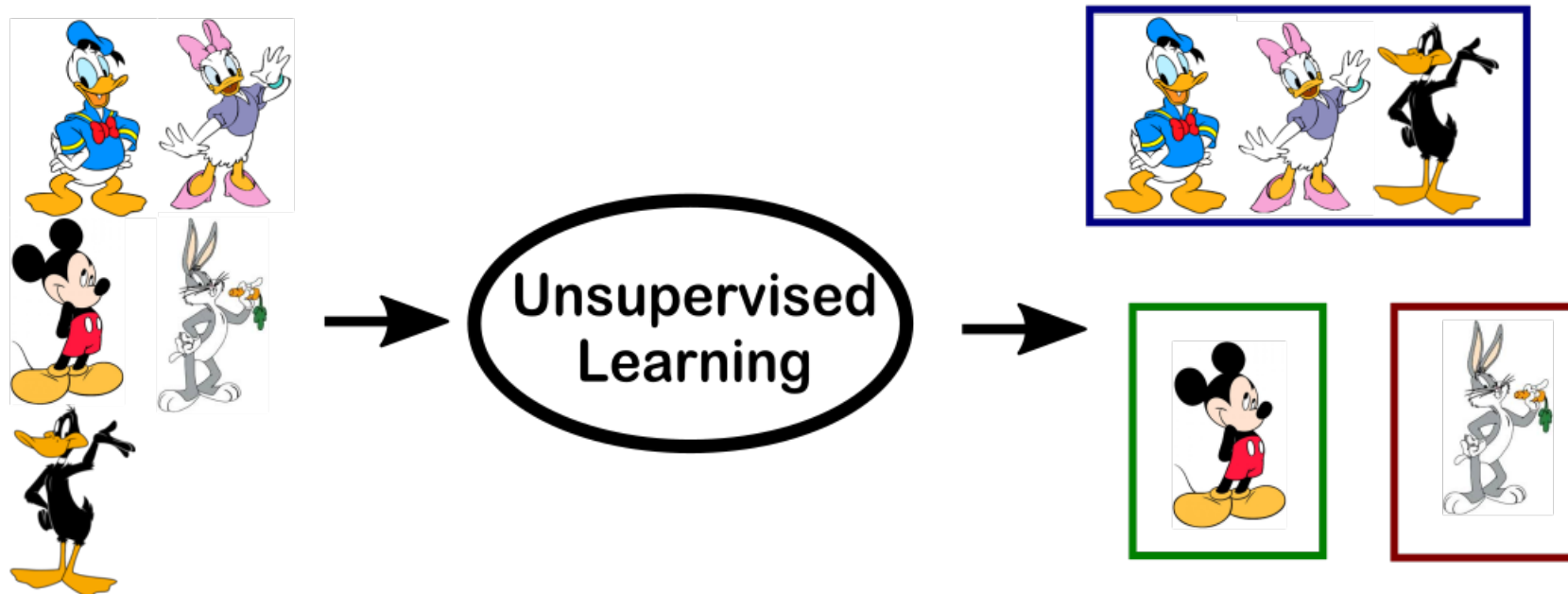
Suppose you had a fruit basket and your task is to arrange the fruit by type. You can group fruits based on any physical character of that particular fruit. For example, if you arrange them according to the color, then the groups will be in a format like:

Red Color Group: Apples & Cherries

Green Color Group: Watermelons & Grapes

If you consider an additional physical character, size, when you group the fruits, you will get results like:

Red Color and Big Size Group: apple.

Green Color and Big Size Group: Watermelons

Red Color and Small Size Group: cherries

Green Color and Small Size Group: grapes

The Effect of Machine Learning on Web Application Development

Understand Customer Behavior

It examines

- Characteristics of individual consumers such as demographics, personality lifestyles and behavioral variables such as usage rates, usage occasion, loyalty, brand advocacy, willingness to provide referrals, repurchase intentions

- how emotions, attitudes and preferences affect buying behavior.

- Influence on customer from other social groups

- The web applications can also use machine learning algorithms to understand customer behavior and boost customer engagement.

- An e-commerce application can use machine learning to monitor and understand customer conversation related to a product. (comment/feedback)

- It can even use the algorithm to know the features and functionality expected by the customers.

Deliver Personalized Content and Information

Facebook already use machine learning algorithm to personalize news-feed of each user.

- It combines predictive analytics and statistical analysis to detect patterns based on the user's data (content read and posts liked by the user).

- personalizes the news-feed of the users based on the detected patterns.

Speedup Product Discovery

- Large companies like Apple, Google, and Microsoft are already using machine learning algorithms to deliver smart search results to each user

- Use the technology to help customers find right products based on their precise needs.

- E-commerce applications, the programmers can use machine learning algorithm to help customers find products faster.

- The e-commerce portal can further use machine learning to make the customers browse through only relevant products.

DATA MINING

| Extraction of data from repositories | Data cleansing and loading into data mining database | Data transformation | Pattern discovery using algorithms such as *clustering, regression* and *classification* | Data visualization and interpretation of results |