What is Machine Learning?

- Machine learning is programming computers to optimize a performance criterion using example data or past experience.

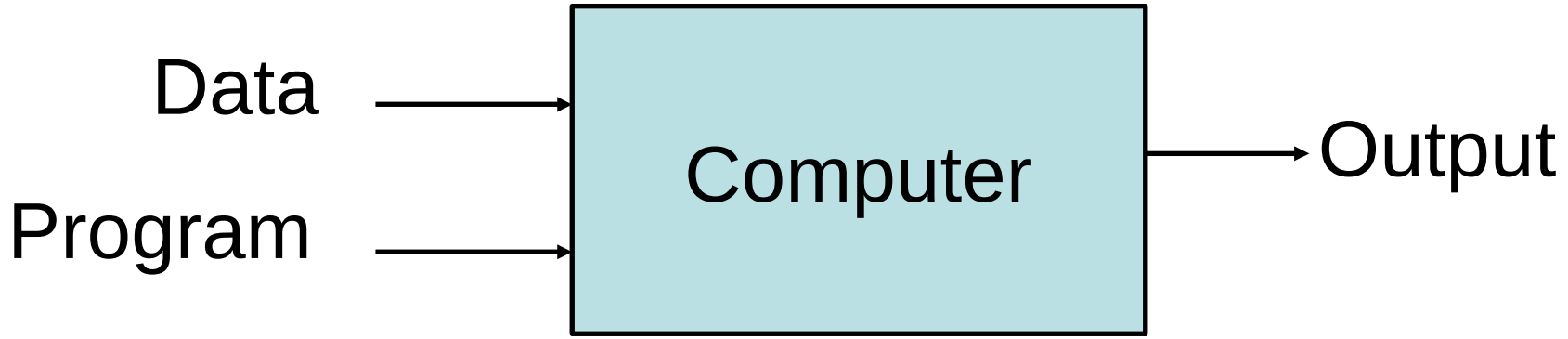
  [-- Ethem Alpaydin]

- The goal of machine learning is to develop methods that can automatically detect patterns in data, and then to use the uncovered patters to predict future data or other outcomes of interest.

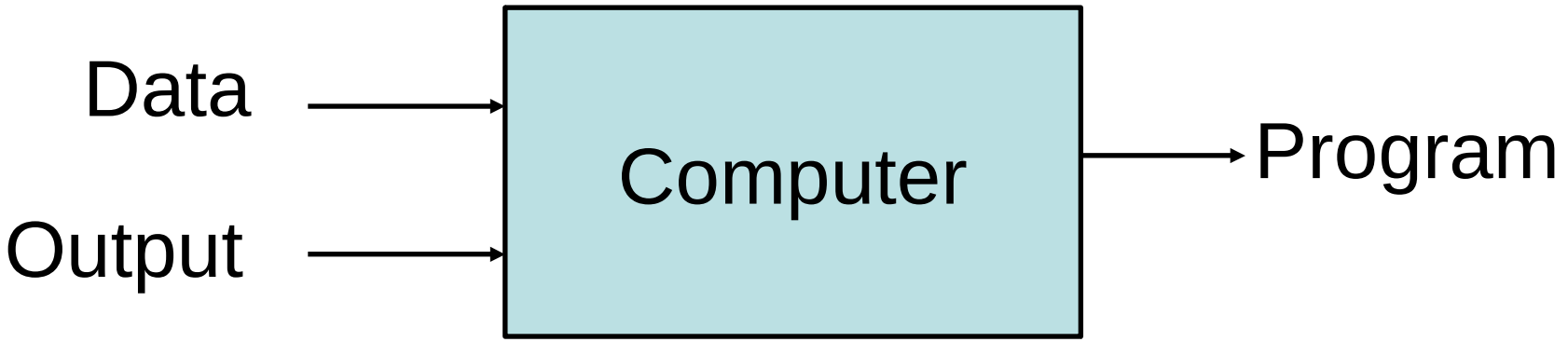
  [-- Kevin P. Murphy]

Based on slide by David Kauchak

Traditional Programming



Machine Learning



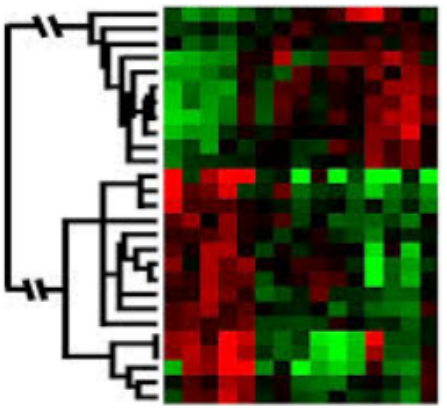
Slide credit: Pedro Domingos

When Do We Use Machine Learning?

ML is useful when:

- Human expertise does not exist (navigating on Mars)
- Humans can’t explain their expertise (speech recognition)
- Models must be customized (personalized medicine)
- Models are based on huge amounts of data (genomics)



Based on slide by Eric Eaton

Classic Example:
It is hard to say what makes a 2

Slide Credit: Geoffrey Hinton

Other examples of tasks best solved by machine learning

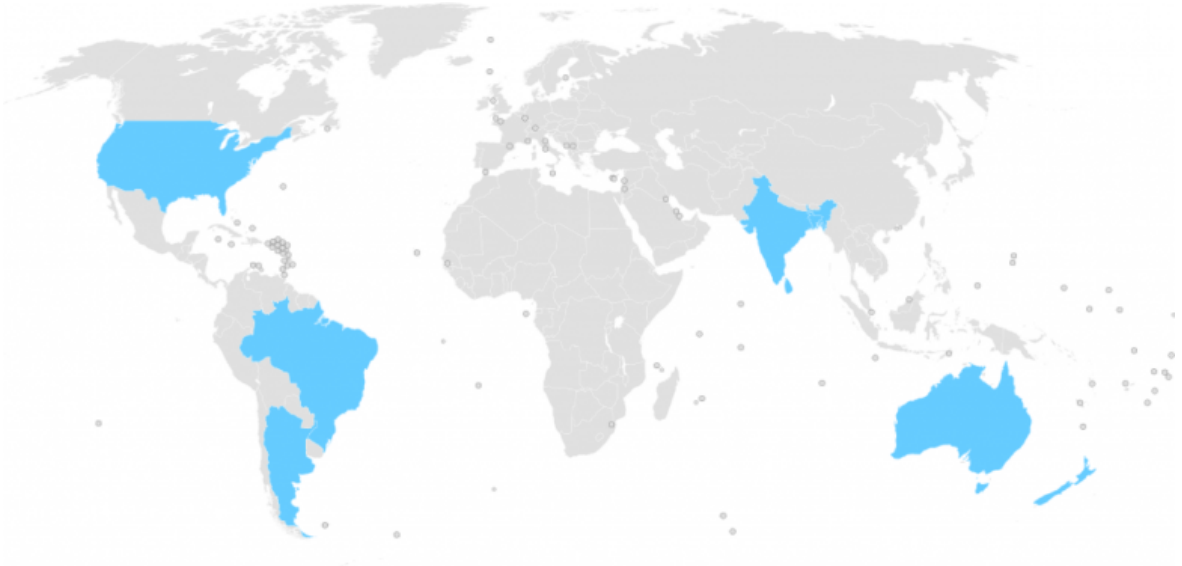
- Recognizing patterns:
  - Objects in real scenes
  - Facial identities or facial expressions
  - Spoken words
- Recognizing anomalies:
  - Unusual sequences of credit card transactions
  - Unusual patterns of sensor readings in a nuclear power plant
- Prediction:
  - Future stock prices or currency exchange rates
  - Which movies will a person like?

Slide Credit: Geoffrey Hinton

This Year's WiDS Datathon

- Focused on **patient health data** from MIT's GOSSIS (Global Open Source Severity of Illness Score) initiative

130,000 hospital Intensive Care Unit (ICU) visits from GOSSIS consortium



kaggle Microsoft

WEST BIG DATA INNOVATION HUB ICME Stanford Institute for Computational & Mathematical Engineering

MathWorks IBM

Berkeley UNIVERSITY OF CALIFORNIA MIT Critical Data

UNIVERSITY OF TORONTO Walmart Labs

This Year’s WiDS Datathon

- Goal is to predict **patient survival** using data from the first 24 hours of intensive care

hospital_death	None	binary	Whether the patient died during this hospitalization
age	Years	numeric	The age of the patient on unit admission
bmi	kilograms/metres^2	string	The body mass index of the person on unit admission
elective_surgery	None	binary	Whether the patient was admitted to the hospital for an elective surgical operation
ethnicity	None	string	The common national or cultural tradition which the person belongs to

Goals for the workshop

- Develop intuition for working with patient data
  - Data exploration and preprocessing
- Understand and apply ML algorithms to data
  - Model representations: logistic regression, decision trees, neural networks, etc.
  - Evaluation metrics and optimization
- Appreciate what makes healthcare different
  - Need to consider **fairness** and **accountability**
  - Small number of samples
  - Lots of missing data
  - etc.

What makes healthcare different?

- Life or death decisions
  - Consider **fairness** and **accountability** of algorithms
- Many of the questions we want to answer are causal
  - Naive use of supervised machine learning is insufficient
- Sometimes a small number of samples
  - Majority of patients are healthy people, survivors, etc.
- Lots of missing data

Based on slides by David Sontag