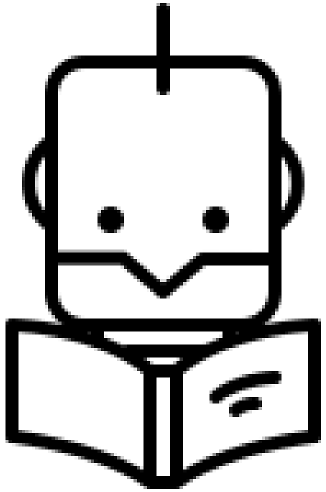
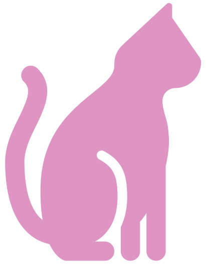
WHAT IS MACHINE LEARNING?



HOW CAN WE SOLVE A SPECIFIC PROBLEM?

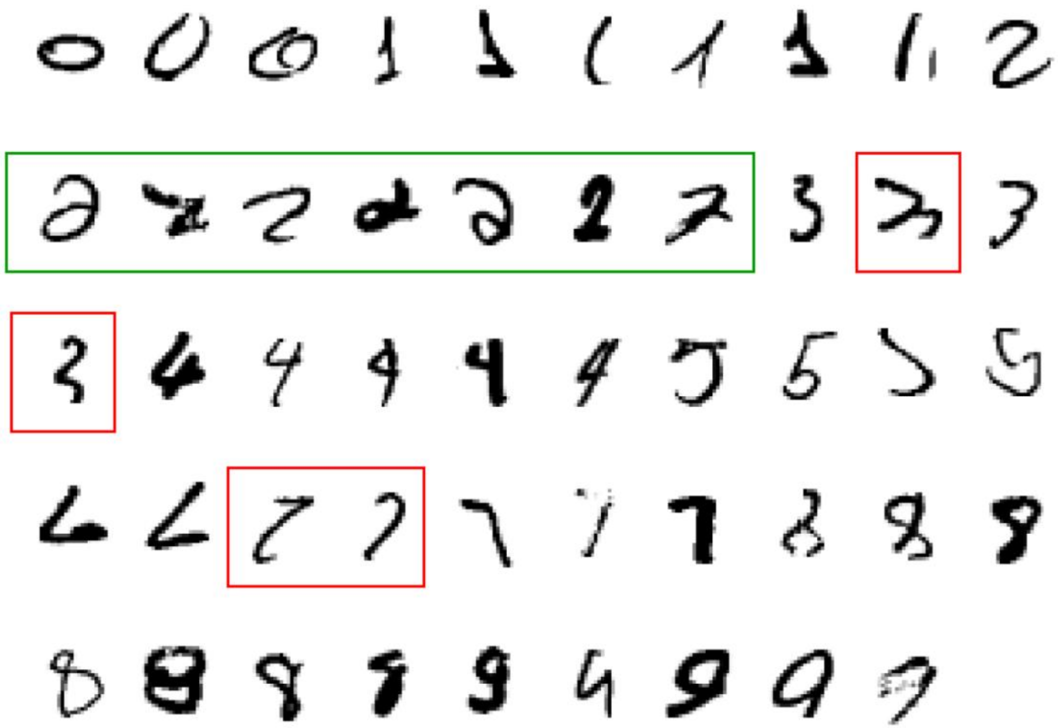
As computer scientist we **write a program** that encodes a set of rules **to solve a problem**



In many cases it is **very difficult to specify those rules**, e.g. detecting cats on images

HOW WOULD A CAT DETECTION PROGRAM LOOK LIKE?

WHAT MAKES A 2 A 2?



WHAT IS MACHINE LEARNING

- Learning systems are not directly programmed to solve a problem, but instead develop own program based on:
  - Examples of how they should behave
  - From Trial-and-error experience trying to solve the problem

- It's different than standard CS:
  - as we want to implement an unknown function, but we only have access to sample input-output pairs (training examples)

- Learning means incorporating information from the training examples into the system

MACHINE LEARNING DEFINITION

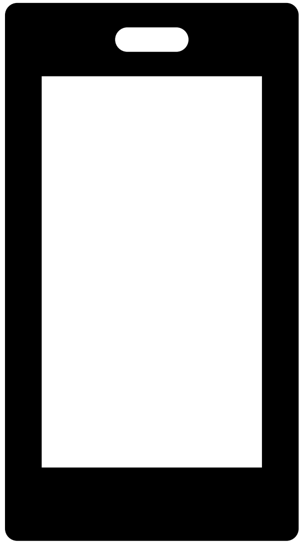
- Arthur Samuel (1959). Machine Learning:
  - Field of study that gives computers the ability to learn without being explicitly programmed.

- Tom Mitchell (1998) Well-posed Learning Problem:
  - A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

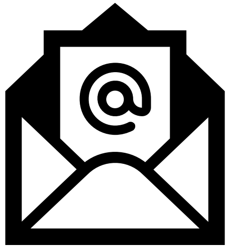
WHERE DO YOU FIND ML?

- Smartphones

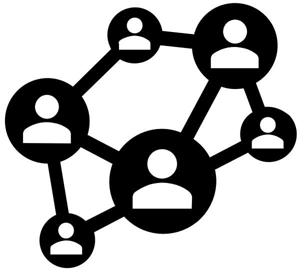
  - Face recognition
  - Voice recognition
  - Personal assistants
  - Fingerprint reader
  - Autocorrect / Text completion



WHERE DO YOU FIND ML? (CONT.)



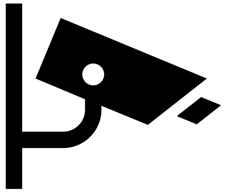
- Email
  - Spam filter



- Social Networks / Social Media
  - Connection recommendations
  - Content recommendations

WHERE DO YOU FIND ML? (CONT.)

- Video surveillance
- Fraud detection
- Many more…



BRIEF HISTORY OF ML

- <1950s – Statistical models
- 1970s – AI Winter: pessimism about machine learning effectiveness
- 1980s – Rediscovery of backpropagation algorithm causes resurgence
- 1990s – Shift from knowledge-driven approach to data-driven approach
  - Neural networks
- 2010 – Deep learning becomes feasible

FAMOUS AI BREAKTHROUGHS

1997 :

IBM Deep Blue chess computer wins against chess grandmaster Garry Kasparov

Not technically a machine learning model, hence arguably not AI

Simulated many moves ahead and made decisions based on defined ruleset

FAMOUS AI BREAKTHROUGHS

2016 :

Google DeepMind’s AlphaGo beats Go Champion Lee Sedol

AlphaGo was a real machine learning model as it improved with experience

Go is by far more computationally complex than chess

WHY NOW?

- Importance of Data
  - With the advent of the Internet it became feasible to collect large amounts of data used for training machine learning models
  - Increased storage capabilities

- Importance of GPUs
  - GPUs today have around 200 times more processors per chip than a CPU
  - GPUs are specialized in parallel operations (addition, multiplication, etc.) on vectors. CPUs would do these operations sequentially and are hence slower

WHAT CAN ML DO?

- Classification
- Pattern Recognition
- Recommender Systems
- Information Retrieval
- Computer Vision
- Robotics
- Playing Games

TERMINOLOGY & NOTATION

- Dataset: A set of data examples, that contain features important to solving a problem

- Instance: A row in the dataset. Sometimes also referred to as *(data) point*, *example* or *observation*. An instance i consists of n feature values $x_j^{(i)}$ for $0 \leq j \leq n$ and, if known, the target outcome $y^{(i)}$.

- Features: Important pieces of data that help us understand a problem. A feature is represented by a column in the dataset and is denoted by the vector x_j.
  $x_j^{(i)}$ represents value of feature j in i^{th} training example.

TERMINOLOGY & NOTATION

Training set of housing prices

Size in feet^2 (x)	Price ($) in 1000's (y)
2104	460
1416	232
1534	315
852	178
…	…

Notation:

m = Number of training examples
x's = "input" variable / feature
y's = "output" variable / "target" variable

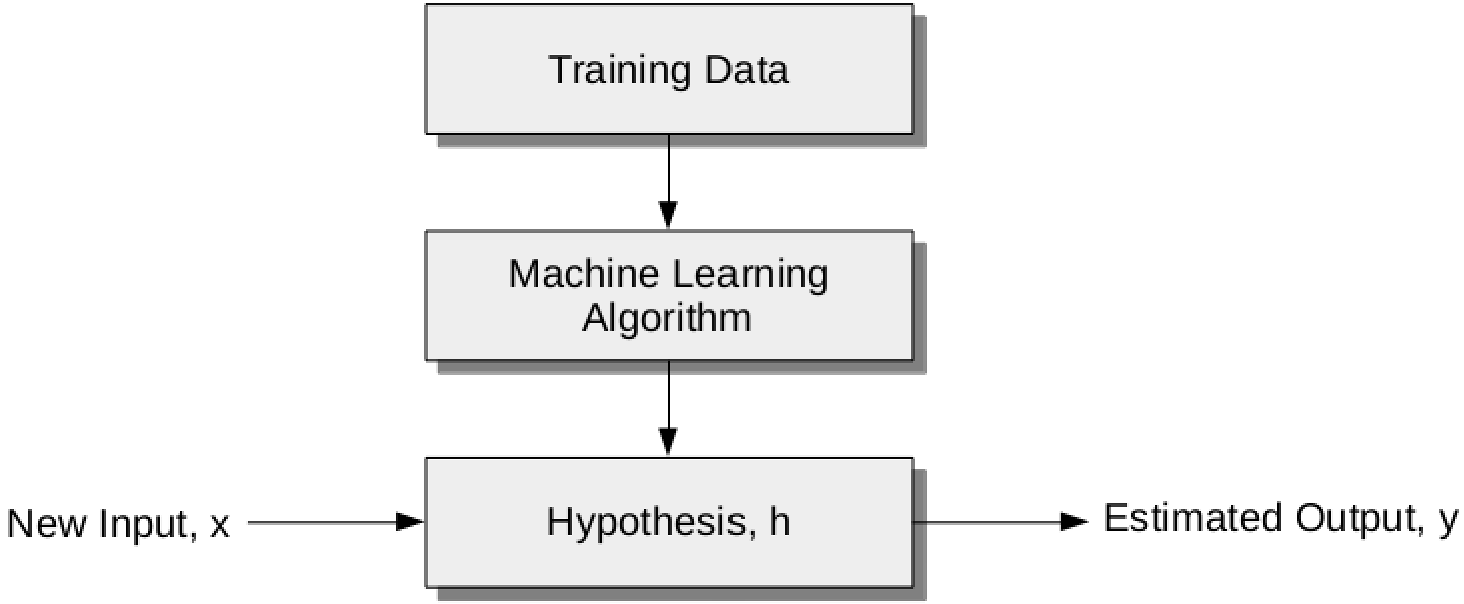
TERMINOLOGY & NOTATION

- Target: is the information the machine learns to predict and is usually denoted by the vector $\mathbf{y}$, or $y^{(i)}$ for a single instance i. In case of a categorical target k denotes the number of categories (labels).

- Prediction: Is what the machine learning model “guesses” what the target value should be, based on the given features of an instance.
  The prediction is often denoted by $\hat{y}$ or $\mathrm{h(x)}$.

TERMINOLOGY & NOTATION

- **Machine Learning Model**: is the learned program that maps inputs to predictions. The model can be a set of **weights** for a linear model or neural network. The target model is often denoted by $\mathbf{h}$.



TERMINOLOGY & NOTATION

- Model parameters:
  Model **parameters** are adapted during the learning phases. These parameters are often referred to as model **weights** and are denoted by Greek letter theta θ.

- Training and Test set:
  A dataset is typically split into a **training set**, which is used in the learning phase to train the machine learning model and a **test set**, which used to evaluate the performance of the model on the machine learning task.

CATEGORIES OF ML

Supervised learning

- Classification (discrete values)
- Regression (continuous values)

Unsupervised learning

- Clustering (discrete values)
- Anomaly detection
- Dimensionality Reduction (continuous values)

Reinforcement learning

- Agent based systems

CATEGORIES OF ML

Supervised learning

- Classification (discrete values)
- Regression (continuous values)

Covered in this course

Unsupervised learning

- Clustering (discrete values)
- Anomaly detection
- Dimensionality Reduction (continuous values)

Reinforcement learning

- Agent based systems

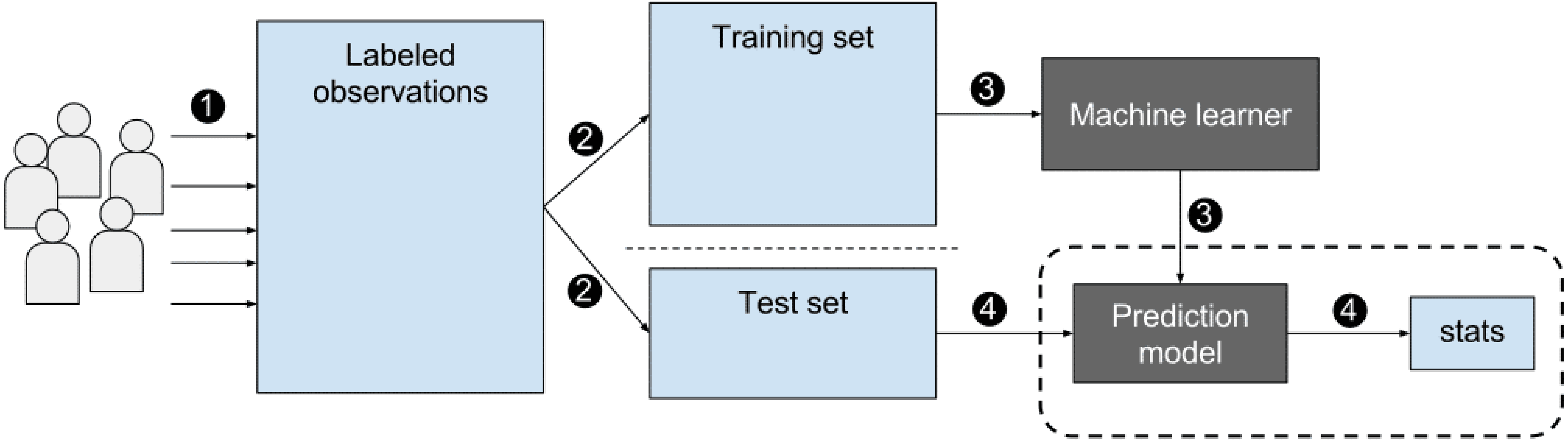
SUPERVISED LEARNING

- In **supervised learning** we have input variables **x** and output variable **y** (sometimes referred to as labels) and the machine learning model learns a mapping function from the input to the output.

- Goal is to approximate the mapping function such that we can use the function to make predictions on **new input** (unseen) data to predict the output variables **y**.

- It is called supervised learning because the process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process.

SUPERVISED LEARNING



SUPERVISED LEARNING

- Supervised learning can be divided in:
  - Classification:
    when the output variable is discrete (category such as “cat” or “dog”)
  - Regression:
    when the output variable is continuous (such as predicting housing prices based on square meters)

UNSUPERVISED LEARNING

- Unsupervised learning does not have information of the corresponding output y of an input variable x during training. The goal of unsupervised learning is to model the underlying structure or distribution of the data.

- Example: Clustering
  - A clustering problem aims to discover inherent groupings in the data, such as grouping customers by purchasing behavior.

ML VS DATA MINING

- Data mining typically using very simple machine learning techniques on very large databases because computers are too slow to do anything more interesting with ten billion examples.

- The term was originally used in a negative sense. Using statistical procedures of looking for all kinds of relationships in the data until finally find one

- Nowadays lines are blurred: many ML problems involve tons of data

ML VS STATISTICS

- ML uses statistical theory to build models
- A lot of ML is rediscovery of things statisticians already knew for a long time; often disguised by differences in terminology
- But the emphasis is different:
    - Good statistics: Clever proff that relatively simple estimation procedure is asymptotically unbiased
    - Good ML: Demo that a complicated algorithm produces impressive results on a specific task
- ML can be viewed as applying computation techniques to statistical problems