



# Machine Learning Basics

Datta Lab presentation @ Princeton University

2023-06-29

Martin Lellep  
[martin.lellep@gmail.com](mailto:martin.lellep@gmail.com)

# Purpose of this talk

Teach you the very basic idea behind  
**Machine Learning (ML)**

# Course materials are available!



<http://lellep.xyz/blog/datta-lab-ml-course.html>

# Questions

- 1) Who has **heard** of ML?
- 2) Who has **used** ML in some way?

# Questions

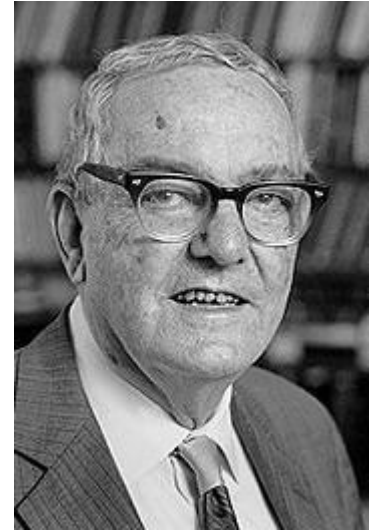


Feel free to ask questions!

# What is ML?

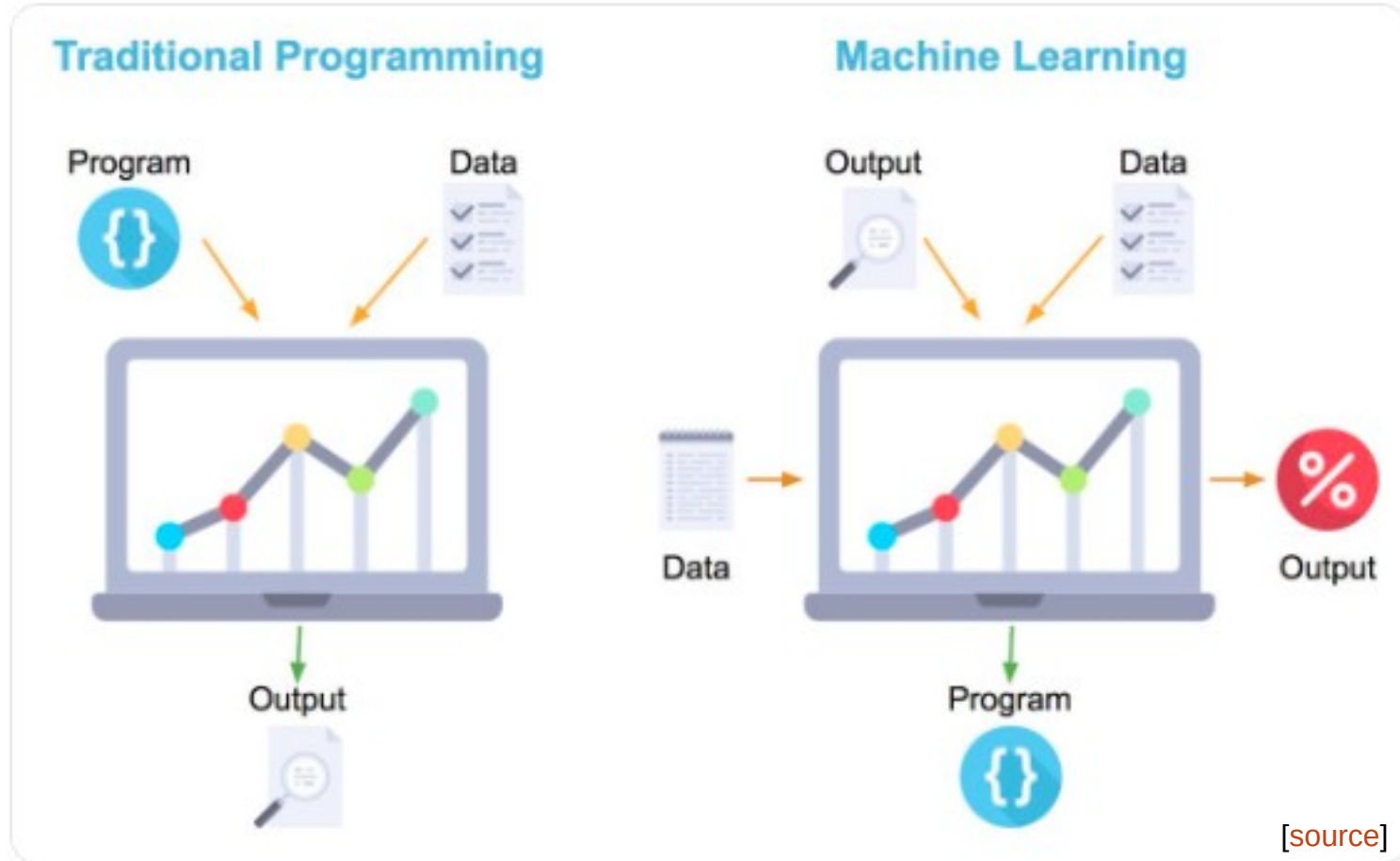
“ML is concerned with computer programs that automatically improve their performance through experience.”

- Herbert Simon



Herbert Simon  
(Turing Award &  
Nobel Prize)

# What is ML?



# ML in daily life

- Most online services
  - Facebook
  - Google
  - Amazon
- Everywhere in daily life
  - Social media
  - Phones
  - Cars

Rucanor Aluminium Baseball Bat  
by Rucanor  
★★★★☆ (29 customer reviews)  
RRP: £24.00  
Price: **£18.40** & **FREE Delivery** in the UK. [Details](#)  
You Save: £2.60 (12%)  
Colour: Silver  
Size: 60  
**In stock.**  
Sold by [HA sports.uk](#) and [Fulfilled by Amazon](#). Gift-wrap available.  
• 60 CM BASEBALL BAT  
• ALUMINIUM Alloy  
[5 new](#) from £14.40

Roll over image to zoom in  
[Share your own customer images](#)

**Frequently Bought Together**

Price For Both: **£28.39**  
[Add both to Basket](#)  
Show availability and delivery details

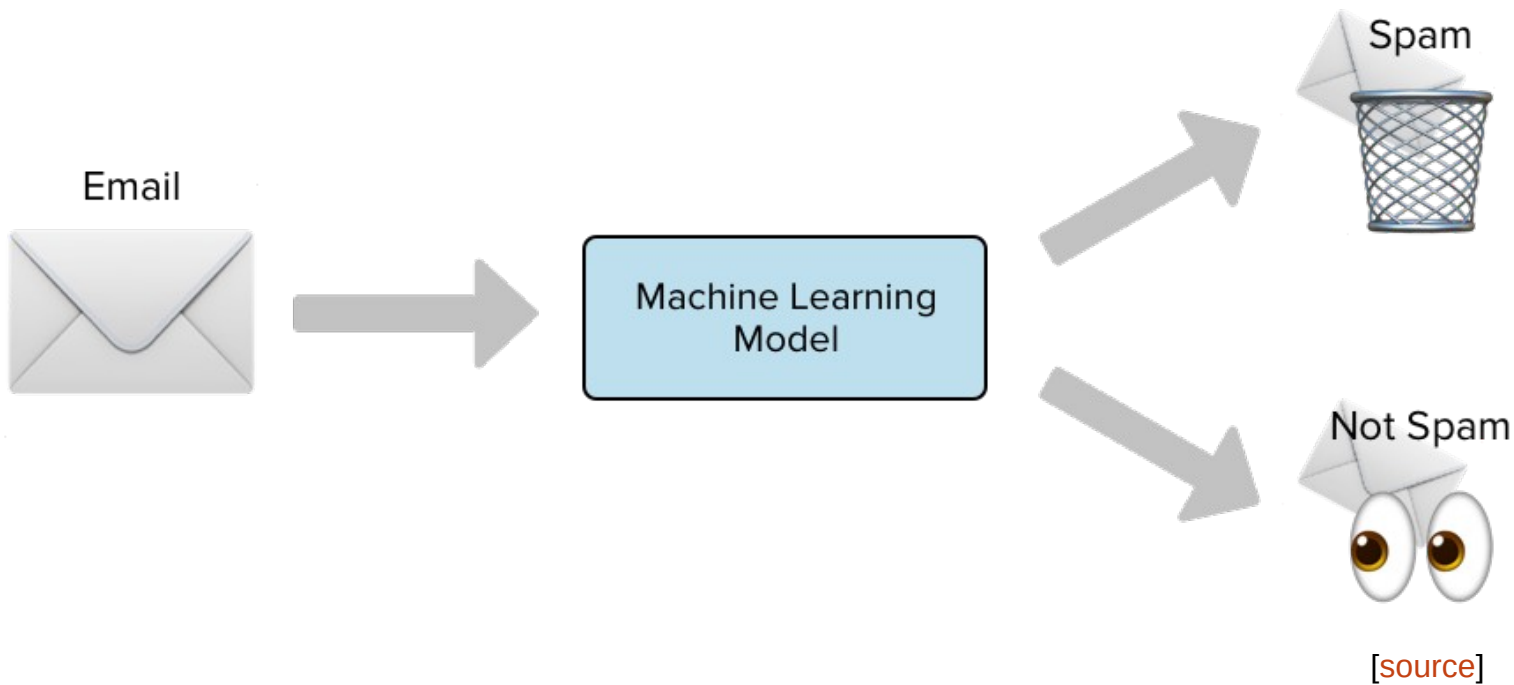
☒ **This item:** Rucanor Aluminium Baseball Bat, Silver - 60 cm £18.40  
☒ **WILSON Official League Individual Baseball** £9.99

**Customers Who Bought This Item Also Bought**

Page 1 of 18



# ML in daily life



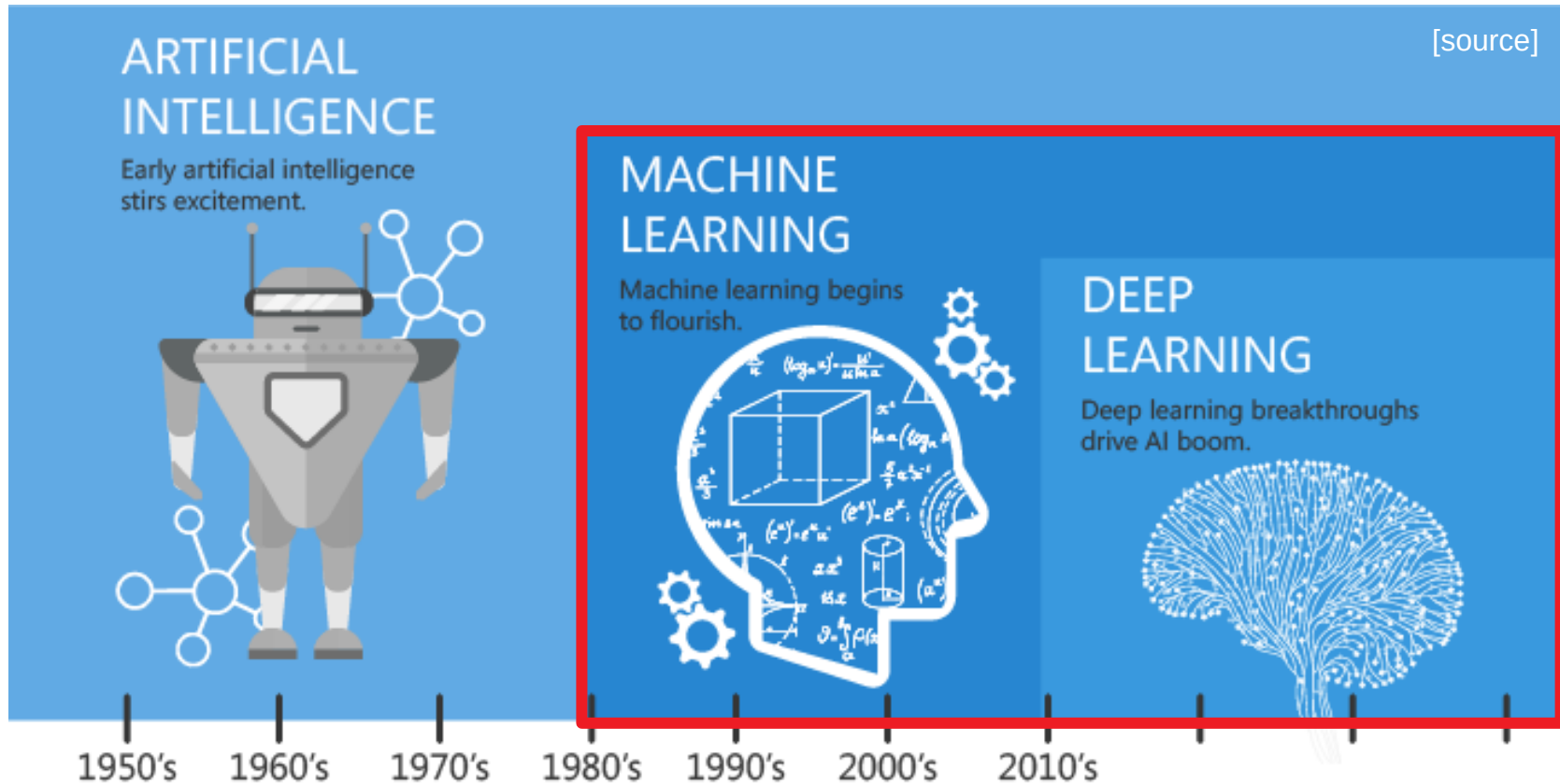
Very popular and in operation for ages already: Spam filters

# Advantages of ML

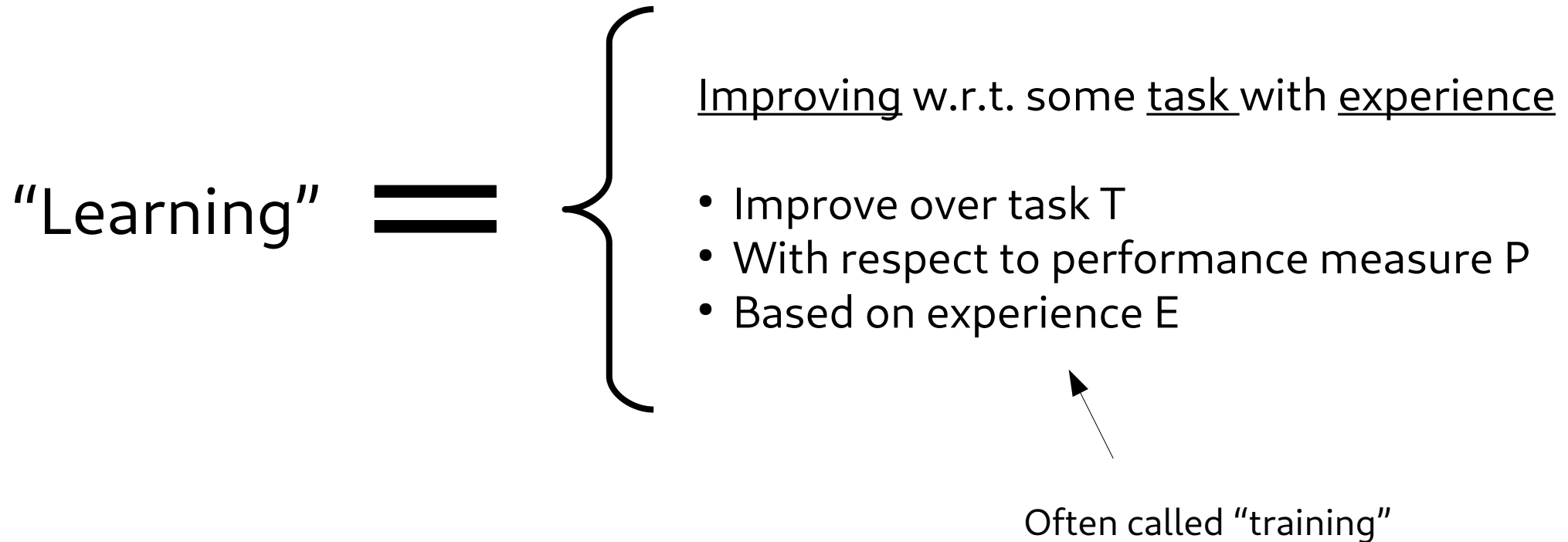
- ML can adapt to data
- ML can solve problems too hard to formulate a traditional algorithm for

Note: There also exist plenty of downsides, too!

# ML and AI: Not to be mixed up



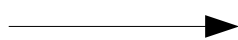
# Concept of ML



# Concept of ML example

- Spam filter
- Here:
  - Task: Identify spam emails
  - Performance: accuracy and errors
  - Experience: Database of emails

Labour-intensive  
and expensive!



Also: Labels of these emails if they were spam or not



# Data: Training, validation and test

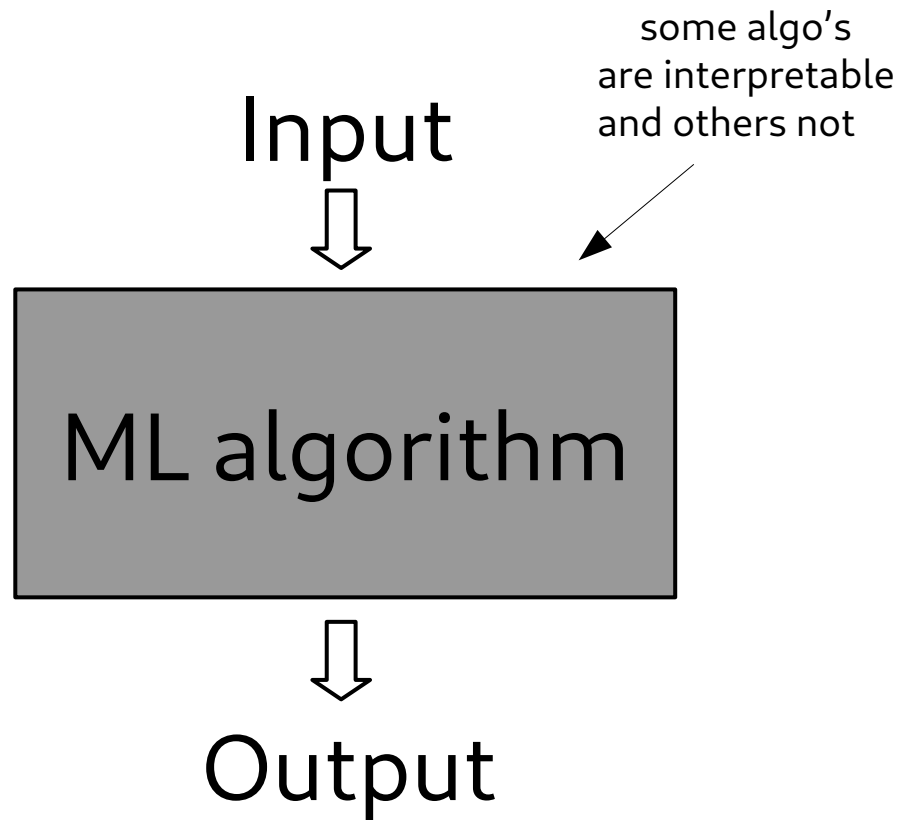
- Experience = data that is shown to ML algorithm
  - Might come in a database ...
  - ... or in a stream of incoming data
- Data is then partitioned
  - Training data: Used to train the ML algorithm
  - Validation data: Check ML algorithm's performance
  - Test data: Task you care about as user

Not used  
for training  
but to inform  
decisions

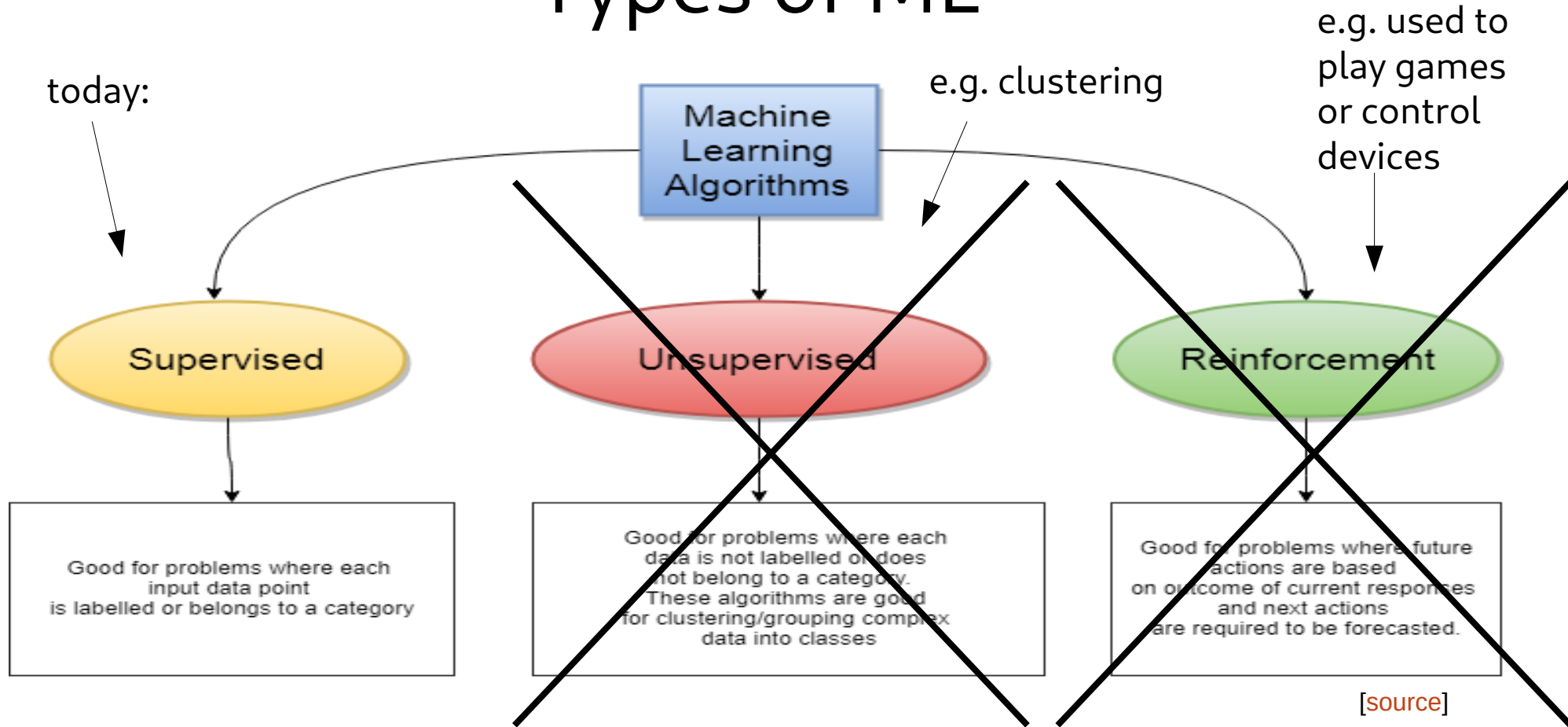


# Samples and features

- Typical setting:
  - Data consists of *samples*
  - Samples consist of *input-output* pairs
- Input is sometimes called *features*



# Types of ML





# Supervised learning

- Given: Training data
  - Input values
  - Desired output
- Goal: For similar input values, generate reasonable output
- Typically: Interpolation, not extrapolation!

# Modi of ML

2 major classes:

- Classification and
  - Regression
- Classification: predict classes of input data
  - Regression: predict numerical values of input data
  - Multiple classes and numerical values possible

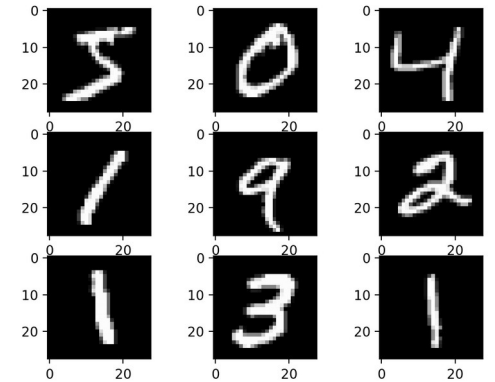
# Modi of ML example

2 major classes:

- Classification and
- Regression

Classification:

- Identify handwritten zip codes
- Make them machine readable



[source]

# Modi of ML example

2 major classes:

- Classification and
- Regression

Regression:

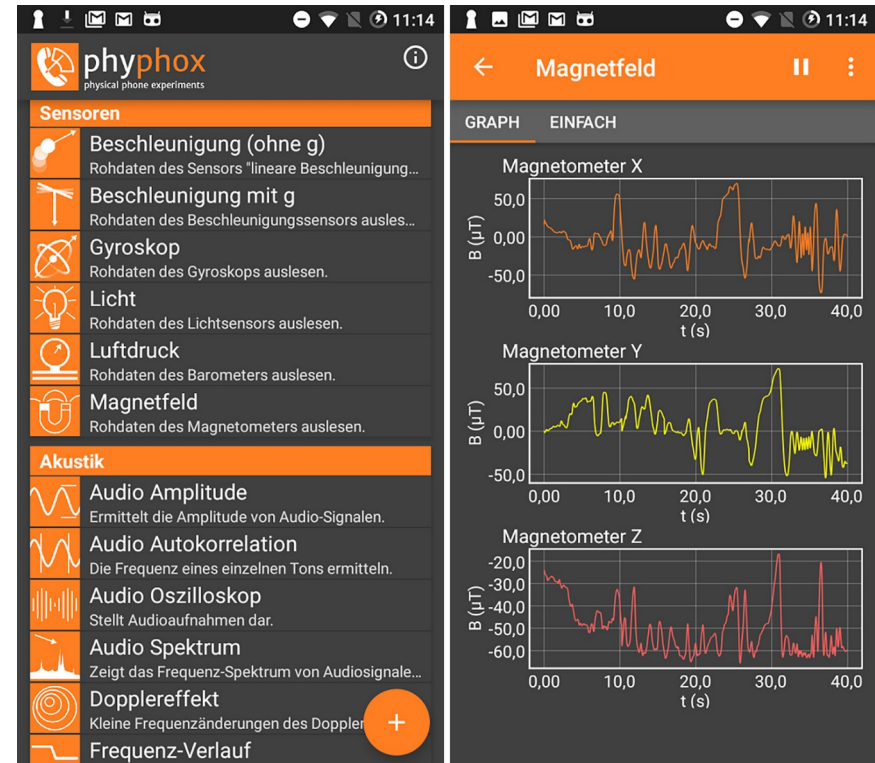
- Given: House ...
  - ... location
  - ... living area
  - ... age
- Predict: Its price

... but wait:  
doesn't that  
sound like  
interpolation?  
**Yes, it does!**



# Live demo: **phyphox**

- Educational app
- Allows access to all your phone's sensors



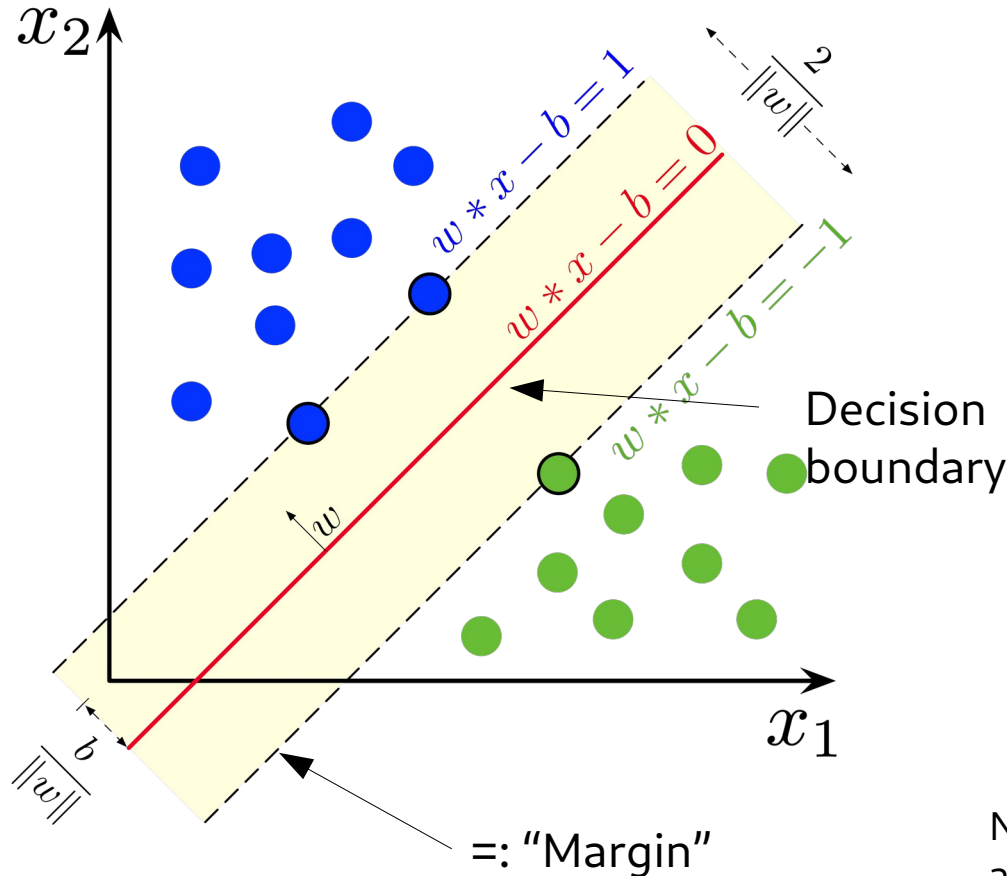
# Live demo: ML with **phyphox**

[Live demo – Part 1 - Preparations]

# Support Vector Machines

- Classical ML algorithm
- Can be used for *classification* and *regression*
- Here: Use it for classification
- Purpose for here:
  - Learn a linear decision boundary of data
  - ... in multi-dimensional space
  - ... learn an optimal one (in some sense)

# Support Vector Machines



[source]

- Decision boundary:  
 $w \cdot x - b = 0$
- Place decision boundary so that margin is maximal

Note: I show here the linear version. But it also works with non-linear features!



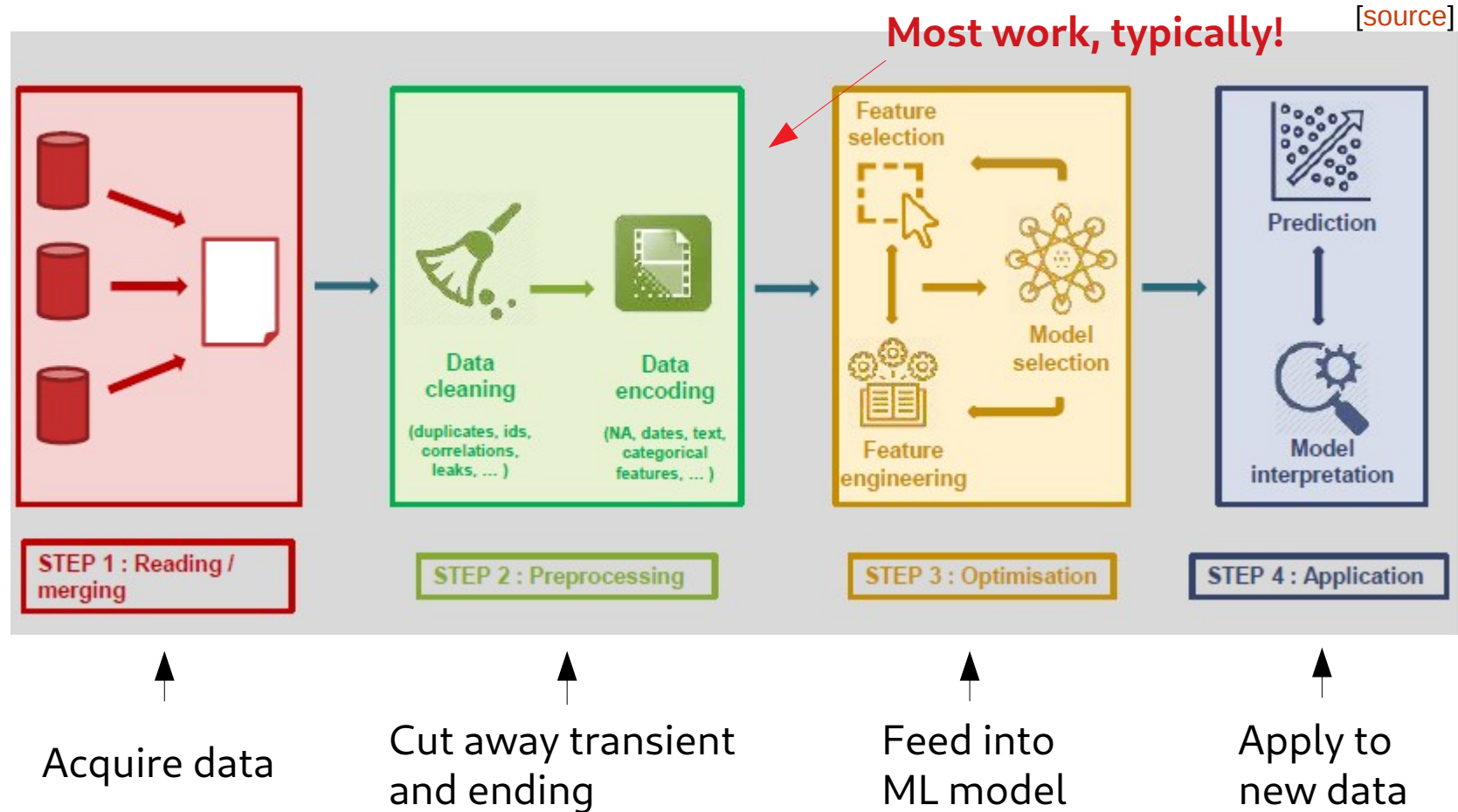
# Example: ML with **phyphox**

[Live demo – Part 2 – ML     ]

# ML pipelines

- ML application consists of multiple steps
- Certain steps repeat for most ML algorithms
- These steps called “pipeline”

# ML pipelines

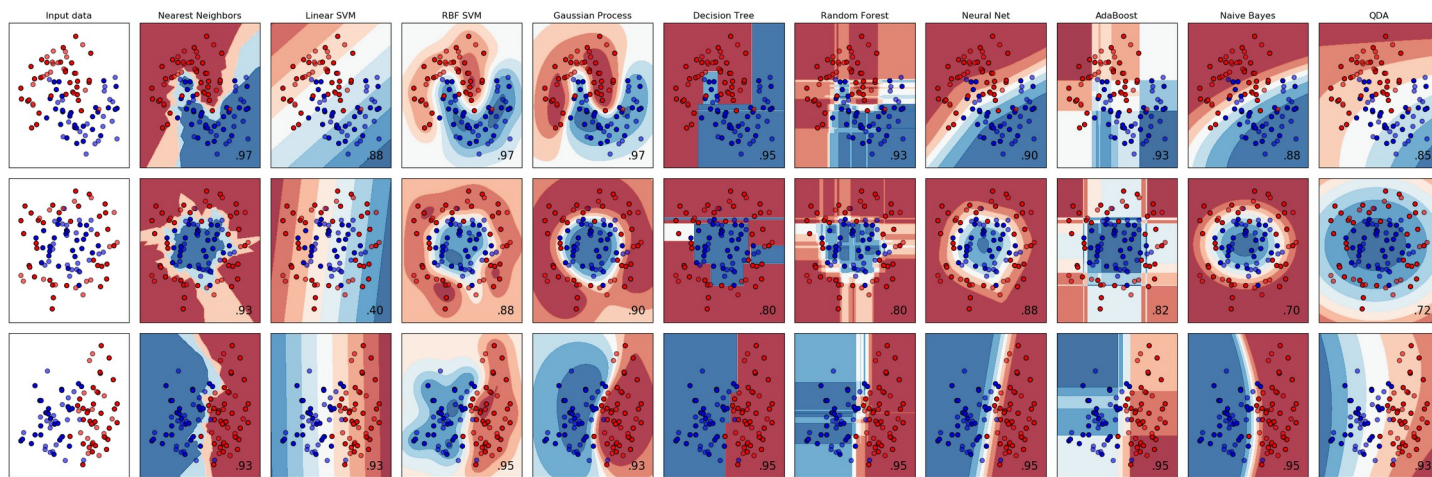


# Frameworks

- Most of the heavy lifting is done by Python packages
- Examples:
  - sklearn
  - Tensorflow
  - PyTorch
  - ... many others!

# Frameworks: Sklearn

- Many ML models available
- Lots of utility functions
- Example: Different classifiers



[source]

# Sources

- ML:
  - <https://de.slideshare.net/liorrokach/introduction-to-machine-learning-13809045>
  - <https://medium.com/fintechexplained/machine-learning-algorithm-comparison-f14ce372b855>
  - <https://medium.com/app-affairs/9-applications-of-machine-learning-from-day-to-day-life-112a47a429d0>
  - <https://towardsdatascience.com/spam-detection-with-logistic-regression-23e3709e522>
- SVM:
  - [https://en.wikipedia.org/wiki/Support-vector\\_machine](https://en.wikipedia.org/wiki/Support-vector_machine)
  - <http://web.mit.edu/6.034/wwwbob/svm-notes-long-08.pdf>
  - <https://blog.aylien.com/support-vector-machines-for-dummies-a-simple/>
  - [https://en.wikipedia.org/wiki/Support-vector\\_machine](https://en.wikipedia.org/wiki/Support-vector_machine)
- Sklearn:
  - [https://scikit-learn.org/stable/auto\\_examples/classification/plot\\_classifier\\_comparison.html](https://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html)

# Get in contact

E-Mail: [martin.llelep@gmail.com](mailto:martin.llelep@gmail.com)



Tech Blog:  
<http://llelep.xyz/blog>



LinkedIn:  
<https://www.linkedin.com/in/martin-llelep-858600152/>

# Questions



Feel free to ask questions!

---

## Next session

Neural networks  
& deep learning

Backup slides



# Unsupervised learning

Examples:

- K-means clustering
- PCA
- Autoencoders
- Generative Neural Networks

# Reinforcement learning

Examples:

- Q-learning
- State Action Reward State Action (SARSA)