

Course 2

Introduction to Machine Learning

Christophe Eloy

Outline

- Practical information
- What is machine learning?
- Different types of learning
- Examples
- Linear models
 - Linear regression with one variable
 - Linear regression with multiple variables

Practical information

- Alternance of classes and tutorials
- Tutorials with Python/Jupyter + standard libraries (numpy, pandas, etc.)
- Assessment with homework + project
- Contact me: christophe.elay@univ-amu.fr

Useful resources

- MOOC Machine Learning, Andrew Ng (Stanford)
- Review paper “A high-bias, low-variance introduction to Machine Learning for physicists”, Pankaj Mehta et al. with Jupyter notebooks
- Lectures on Learning from Data, Yaser Abu-Mostafa (Caltech) on iTunesU and YouTube

What is Machine Learning?

- Two definitions

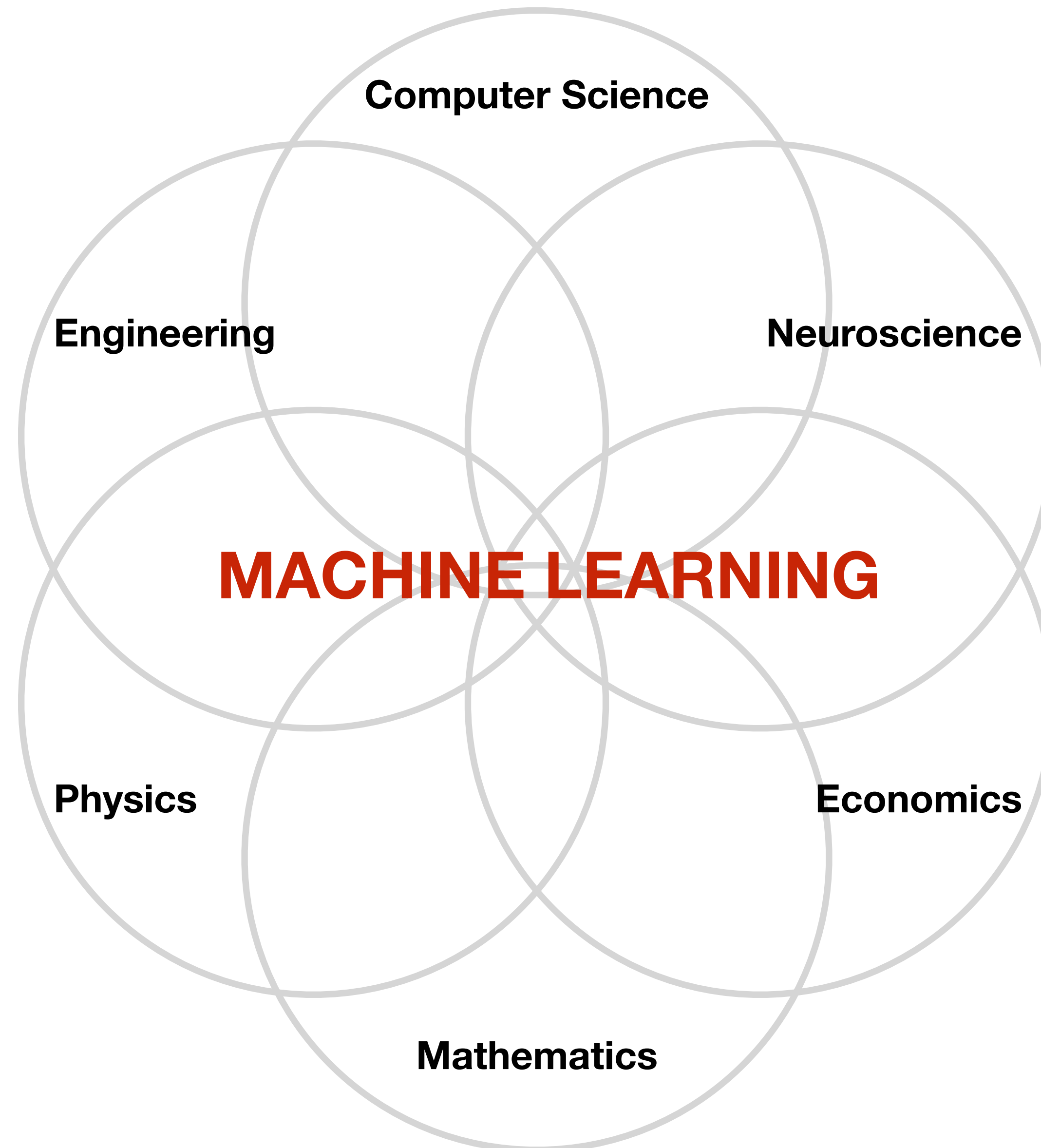
“The field of study that gives computers the ability to learn without being explicitly programmed.”

—Arthur Samuel

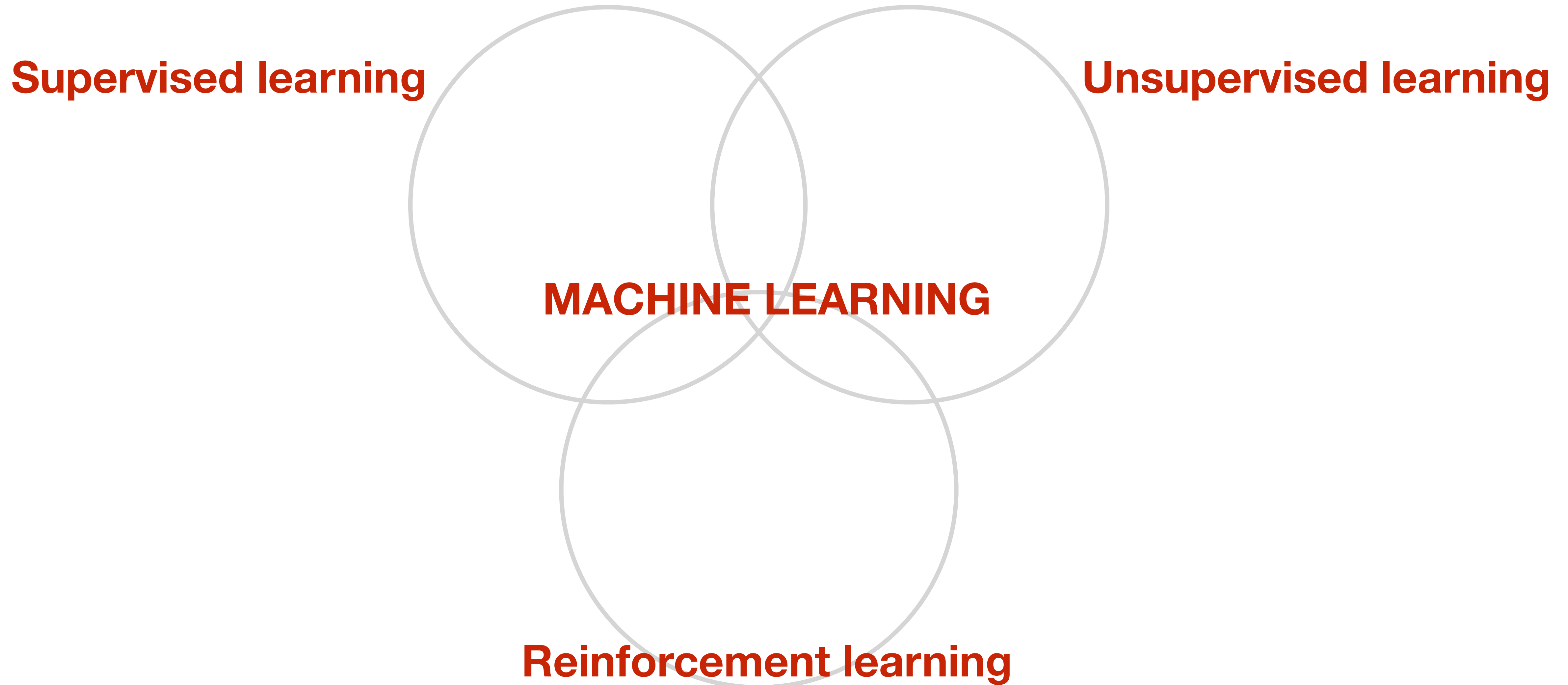
"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E ."

—Tom Mitchell

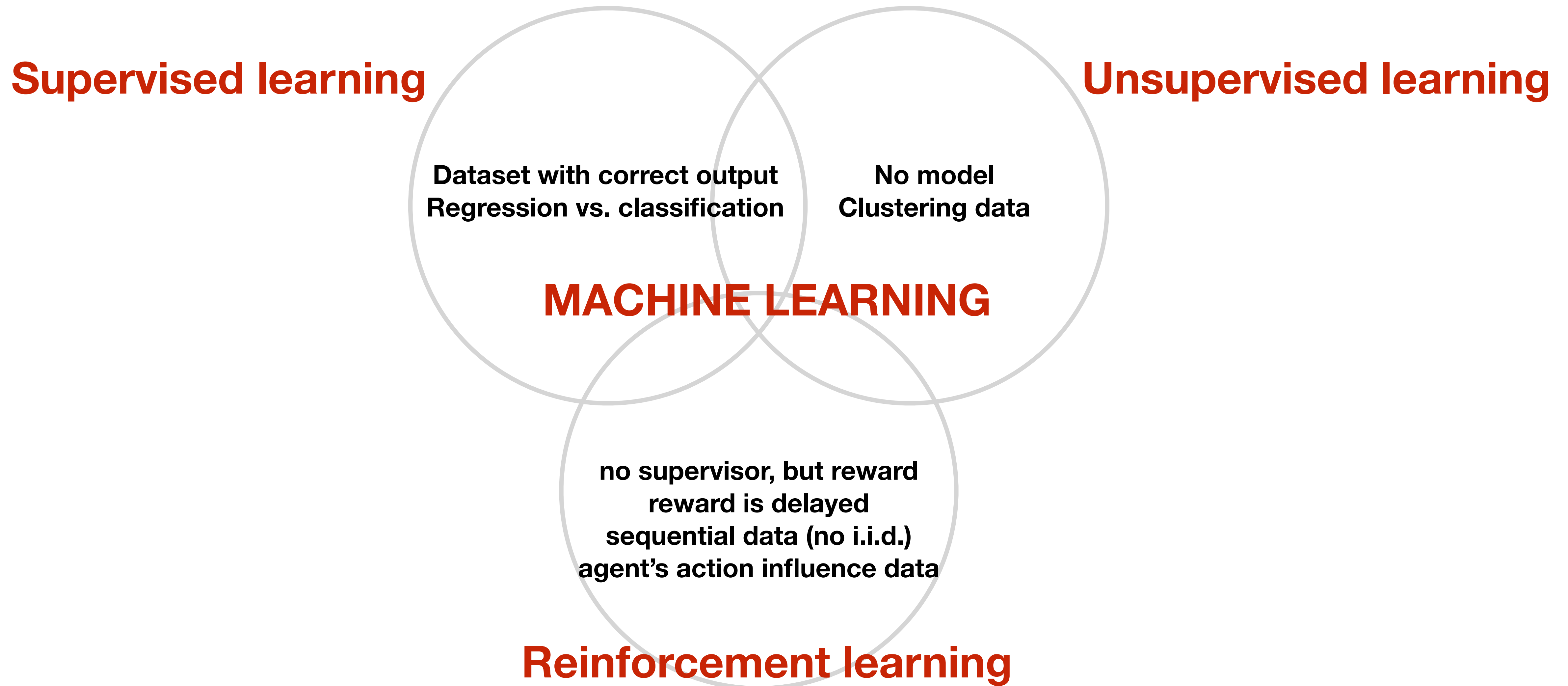
At the crossroad of many fields



Branches of Machine Learning



Branches of Machine Learning



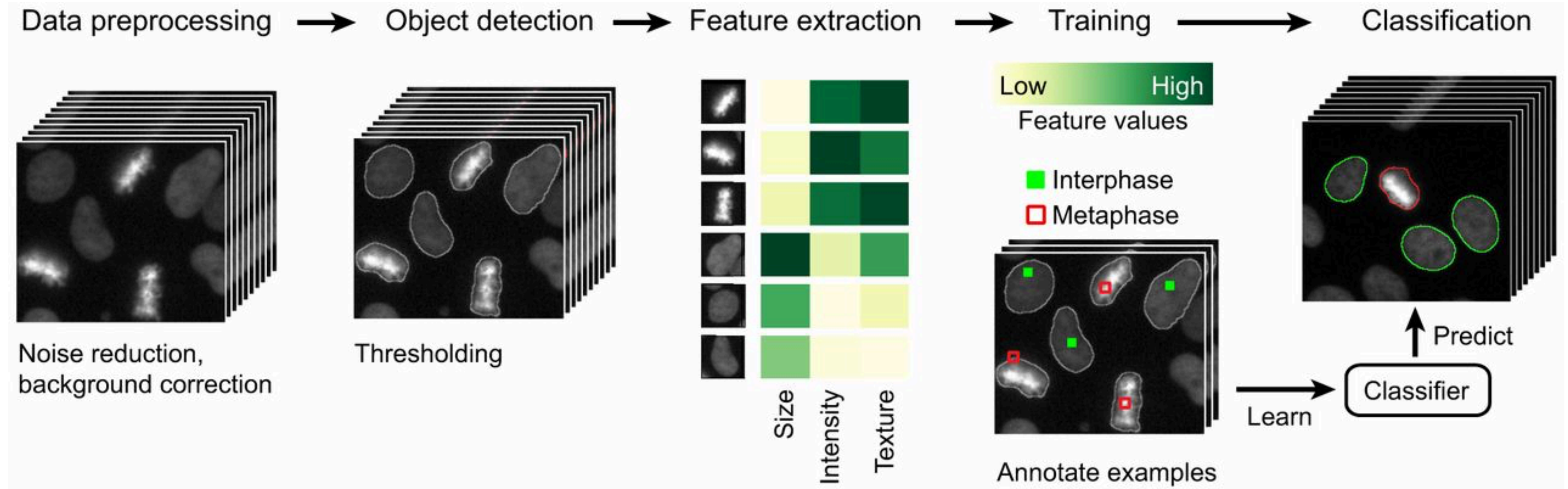
Examples

- Classify emails as spam or not
- Predict the success of a movie
- Diagnose from a list of symptoms
- Drive a car autonomously
- Translate a text
- Find groups in a social network
- Defeat the world champion of Go

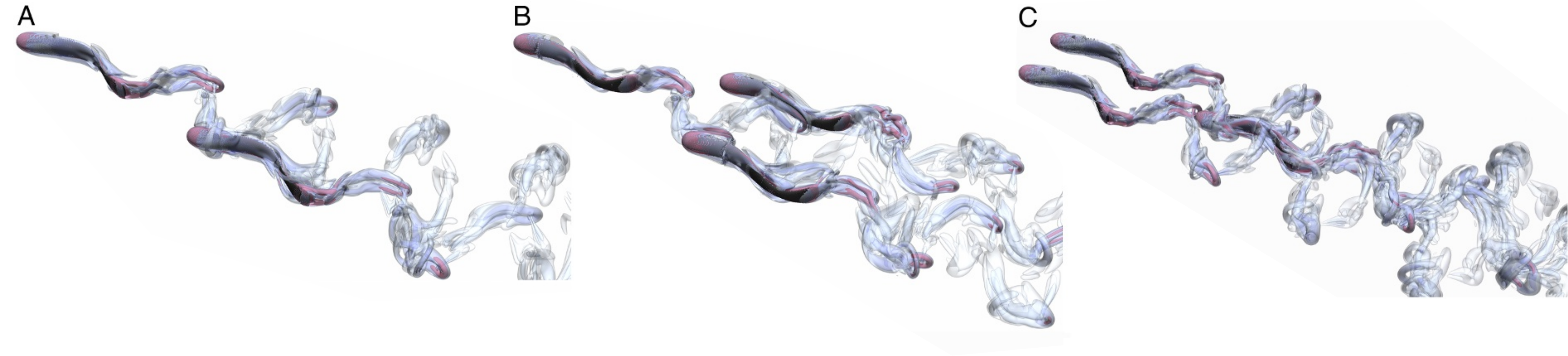
Examples

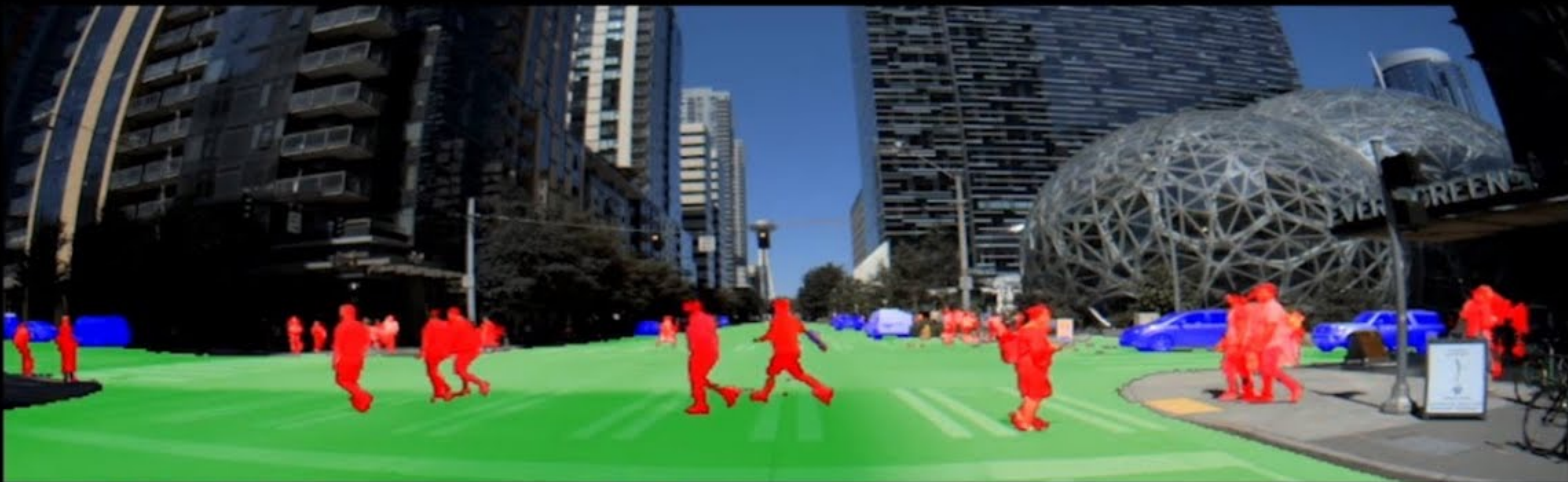
- Classify emails as spam or not Classification (supervised learning)
- Predict the success of a movie Regression (supervised learning)
- Diagnose from a list of symptoms Classification (supervised learning)
- Drive a car autonomously Reinforcement learning
- Translate a text Reinforcement learning
- Find groups in a social network Unsupervised learning
- Defeat the world champion of Go Reinforcement learning

Recognize and classify phenotypes



Learning to exploit vortices







ima...

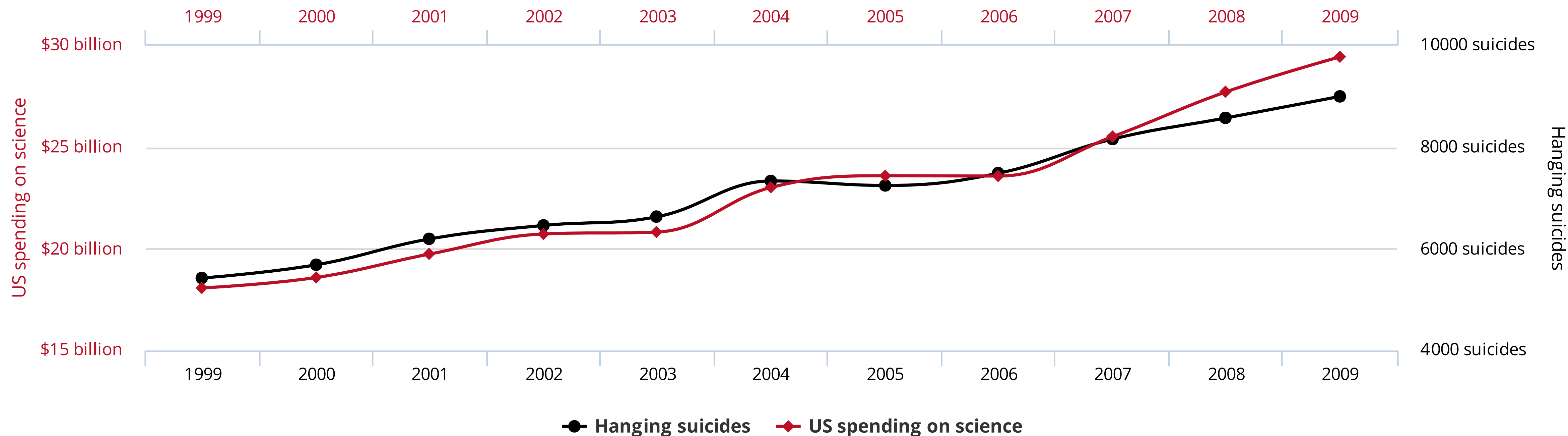


021 3 1

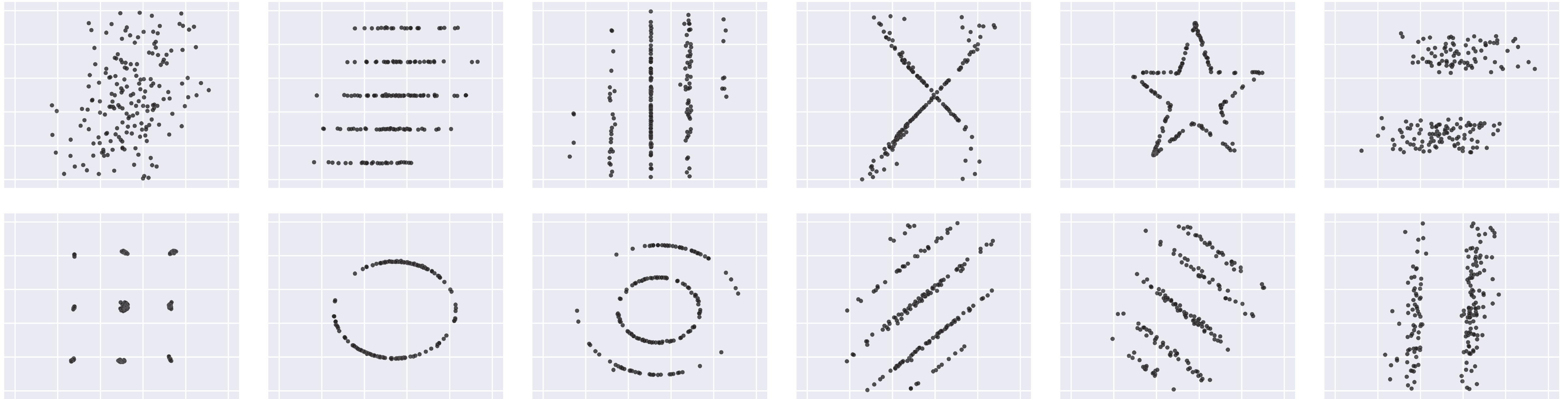


Beware of spurious correlations!

US spending on science, space, and technology
correlates with
Suicides by hanging, strangulation and suffocation



Plot your data!



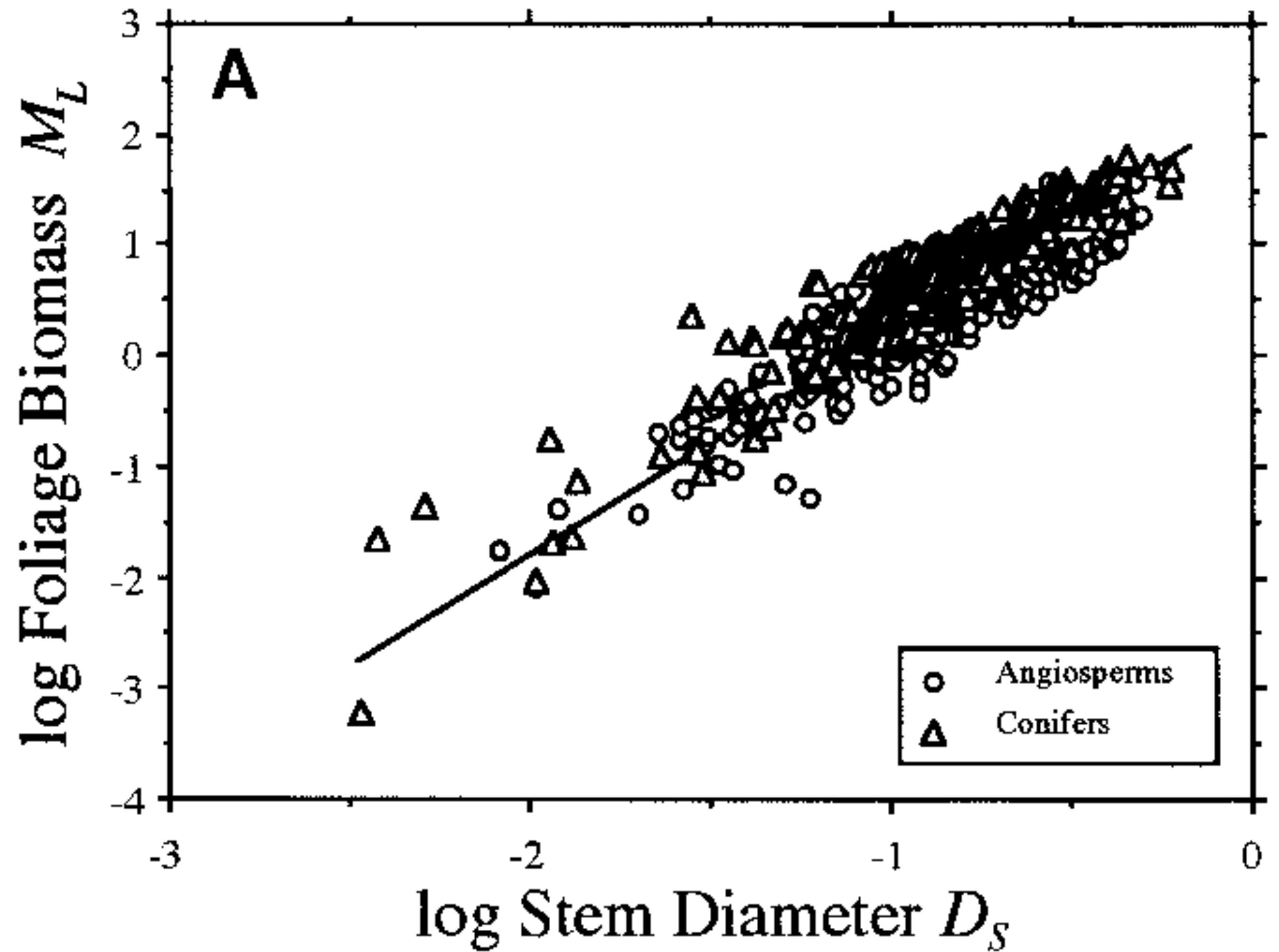
All data have same means, same std, same Pearson's correlation coefficient r

When to use Machine Learning?

- There is a “pattern”
- Relationships are not tractable mathematically
- There is (a lot of) data

Univariate linear regression

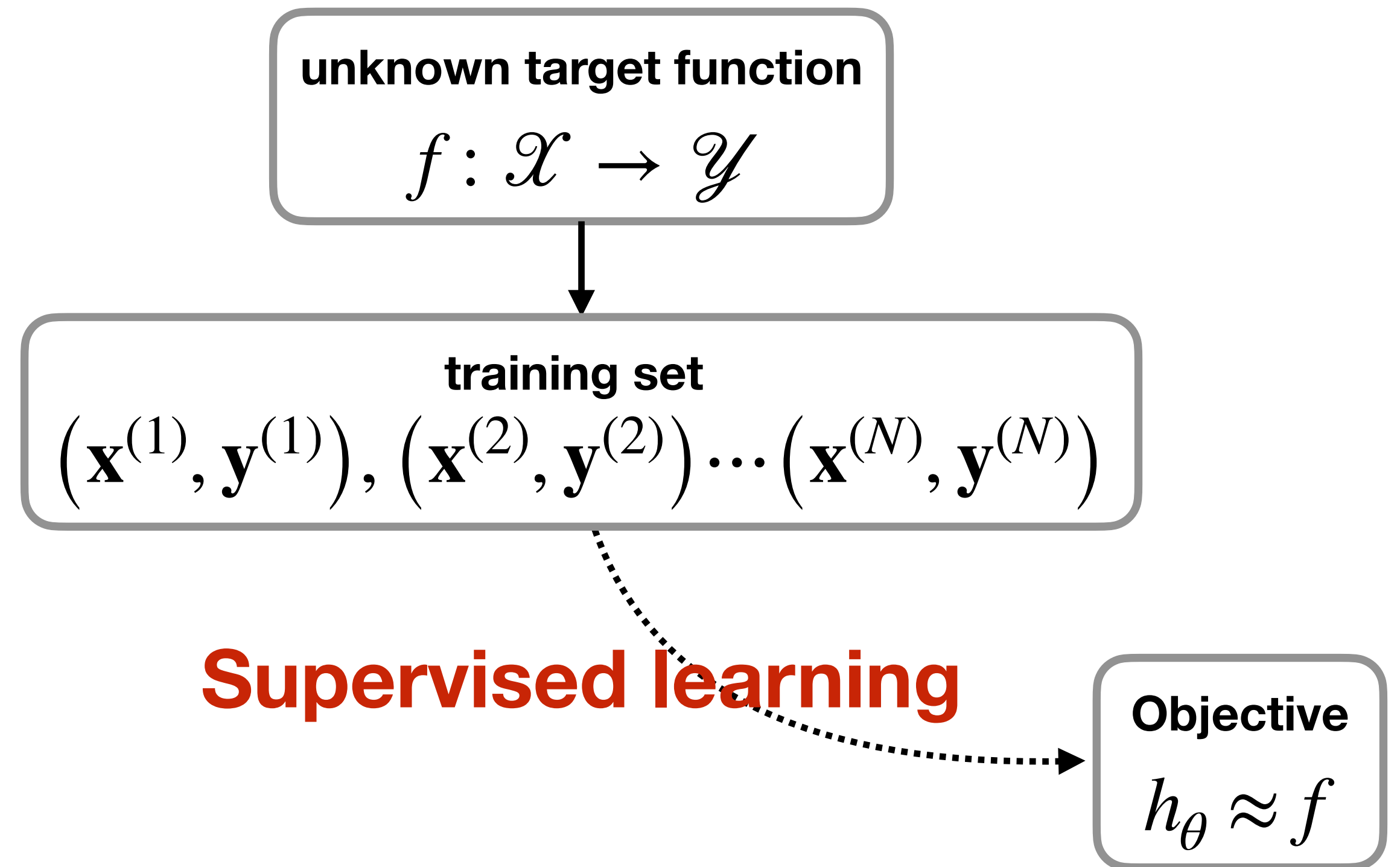
Simple working example



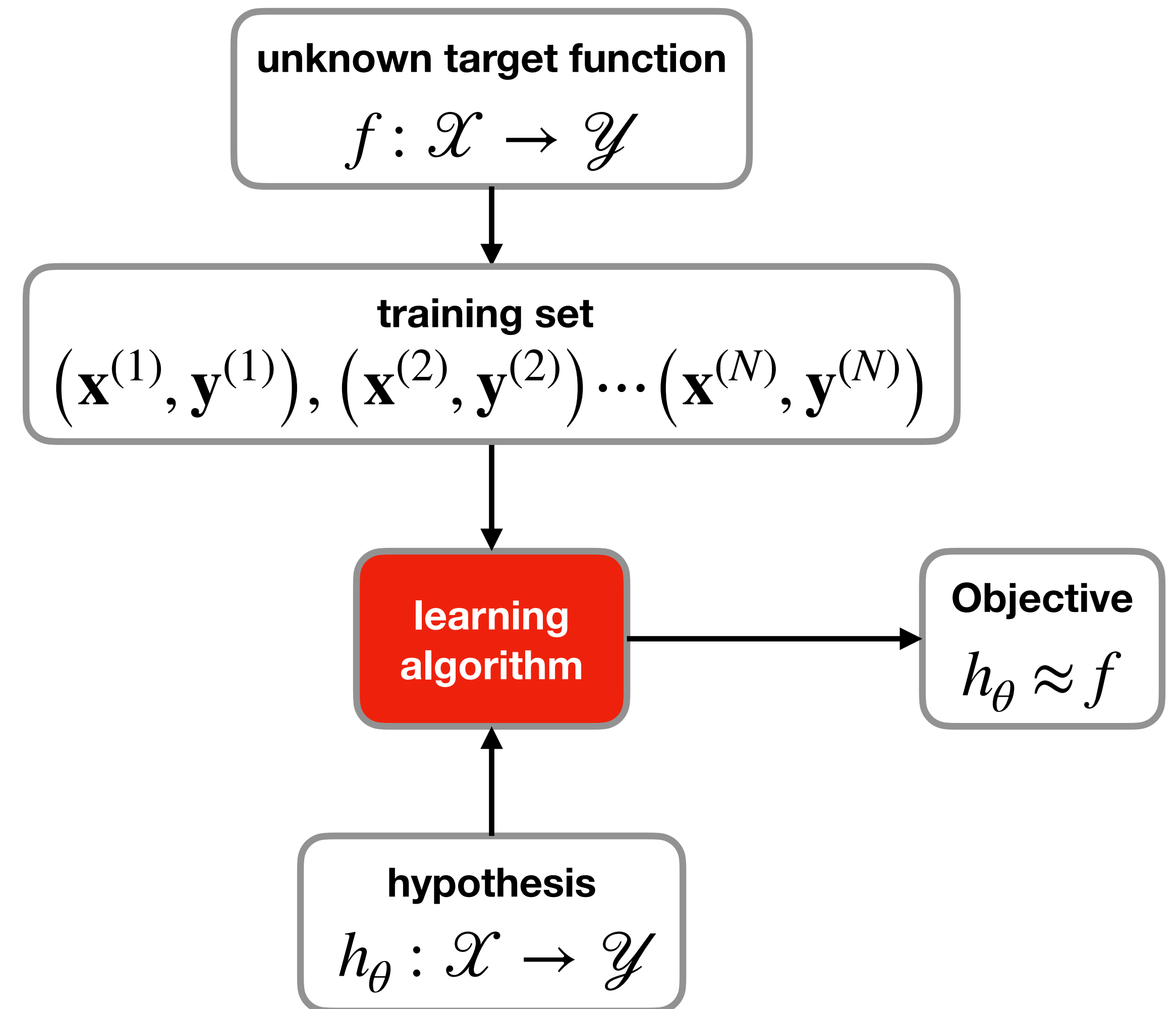
Formalization of learning

- Input: \mathbf{x} (trunk diameter)
- Output: \mathbf{y} (leaf mass)
- Target function: $f: \mathcal{X} \rightarrow \mathcal{Y}$ (the relationship we are looking for)
- Data: $(\mathbf{x}^{(1)}, \mathbf{y}^{(1)}) \cdots (\mathbf{x}^{(N)}, \mathbf{y}^{(N)})$ (to be split into training and testing data)
- Hypothesis: $h_{\theta}: \mathcal{X} \rightarrow \mathcal{Y}$ (the set of functions parametrized by θ)

Learning components

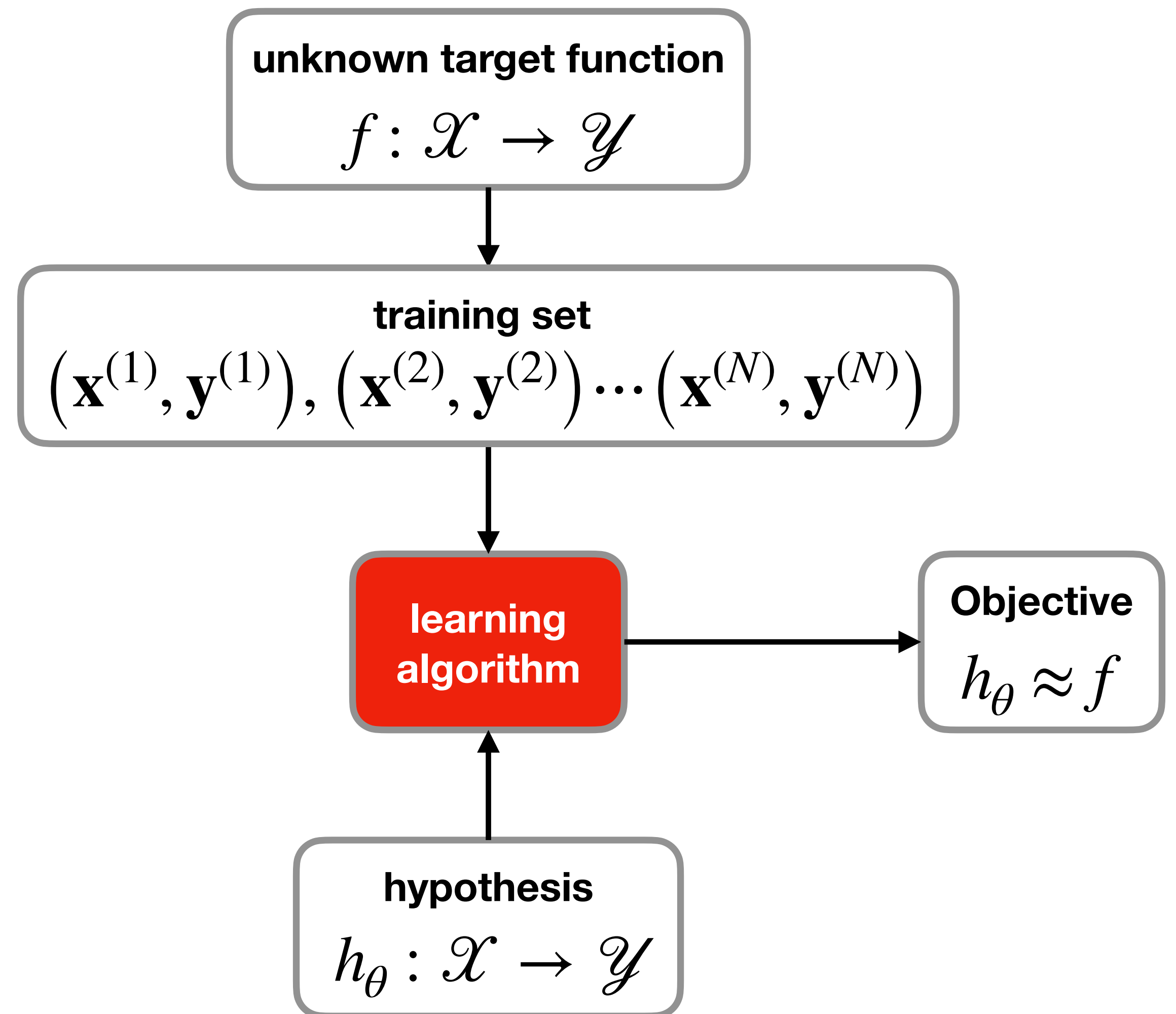


Learning components



Learning components

- Two components of the learning problem
 - Hypothesis: $h_{\theta} \in \mathcal{H}$
 - Learning algorithm
 - Iterative procedure
 - Cost function
- These two components form the **learning model**



Univariate linear regression

- Hypothesis: $h_{\theta}(x) = \theta_0 + \theta_1 x$
- Parameters: $\theta = [\theta_0, \theta_1]^T$
- Cost function (least squares): $J(\theta_0, \theta_1) = \frac{1}{2N} \sum_{i=1}^N (h_{\theta}(x^{(i)}) - y^{(i)})^2$
- Goal: find $\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$

Gradient descent algorithm

- Iterative procedure

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

$$\theta := \theta - \alpha \nabla_{\theta} J(\theta)$$

$$\frac{\partial J}{\partial \theta_0} = \frac{1}{N} \sum_{i=1}^N (h_{\theta}(x^{(i)}) - y^{(i)})$$

$$\frac{\partial J}{\partial \theta_1} = \frac{1}{N} \sum_{i=1}^N (h_{\theta}(x^{(i)}) - y^{(i)}) x^{(i)}$$

- Learning parameter: α
- Full batch vs. stochastic vs. mini-batch gradient descent