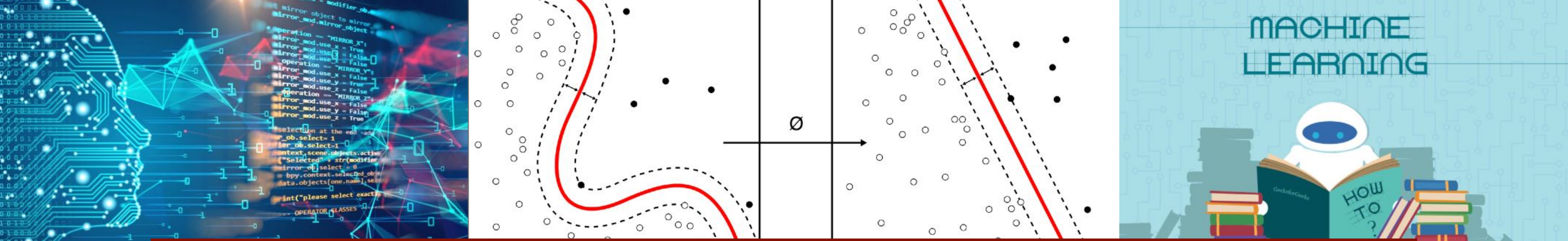


Introduction to Machine Learning

SCP8084699 - LT Informatica

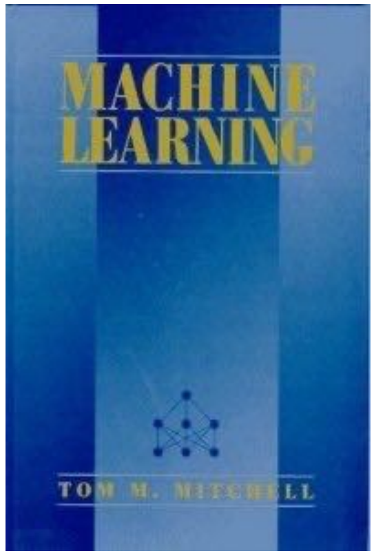
Supervised Learning, Bias-Variance, Linear Regression

Prof. Lamberto Ballan

DIPARTIMENTO
MATEMATICA
Department of Mathematics "Tullio Levi-Civita"

What is Machine Learning?

- "*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 provides a more modern definition (1998)
  - Example: playing checkers
    - Experience *E* = the experience of playing many games of checkers
    - Task *T* = the task of playing checkers
    - Performance measure *P* = percent of games won against opponents



What is Machine Learning?



- **Learning algorithm**: an algorithm that is able to learn from data
- Three main ingredients:
  - The Task
  - The Performance Measure
  - The Experience

The Task

- The task is defined by the problem we want to tackle and the desired output

- Example: classification

Apple

Orange

Mango

The Performance Measure

- How good is the machine learning system?
  - We need to measure its performance, i.e. how good is the function/model returned by the learning algorithm
- The performance measures depends on the task
  - Example (classification): *accuracy* is the proportion of examples for which the model gives the correct output

O:	mango	apple	orange	mango	orange	apple	orange	orange	apple	orange
	1	1	1	1	1	0	0	1	1	0

accuracy = 70%

The Experience

- The experience is provided by the available data
- Which kind of data?
  - Real-valued features, discrete features, …
- How do we get data
  - Obtained once for all (batch), acquired incrementally by interacting with the environment (online learning)
- How can data be used?
  - Learning paradigms

Main Learning Paradigms

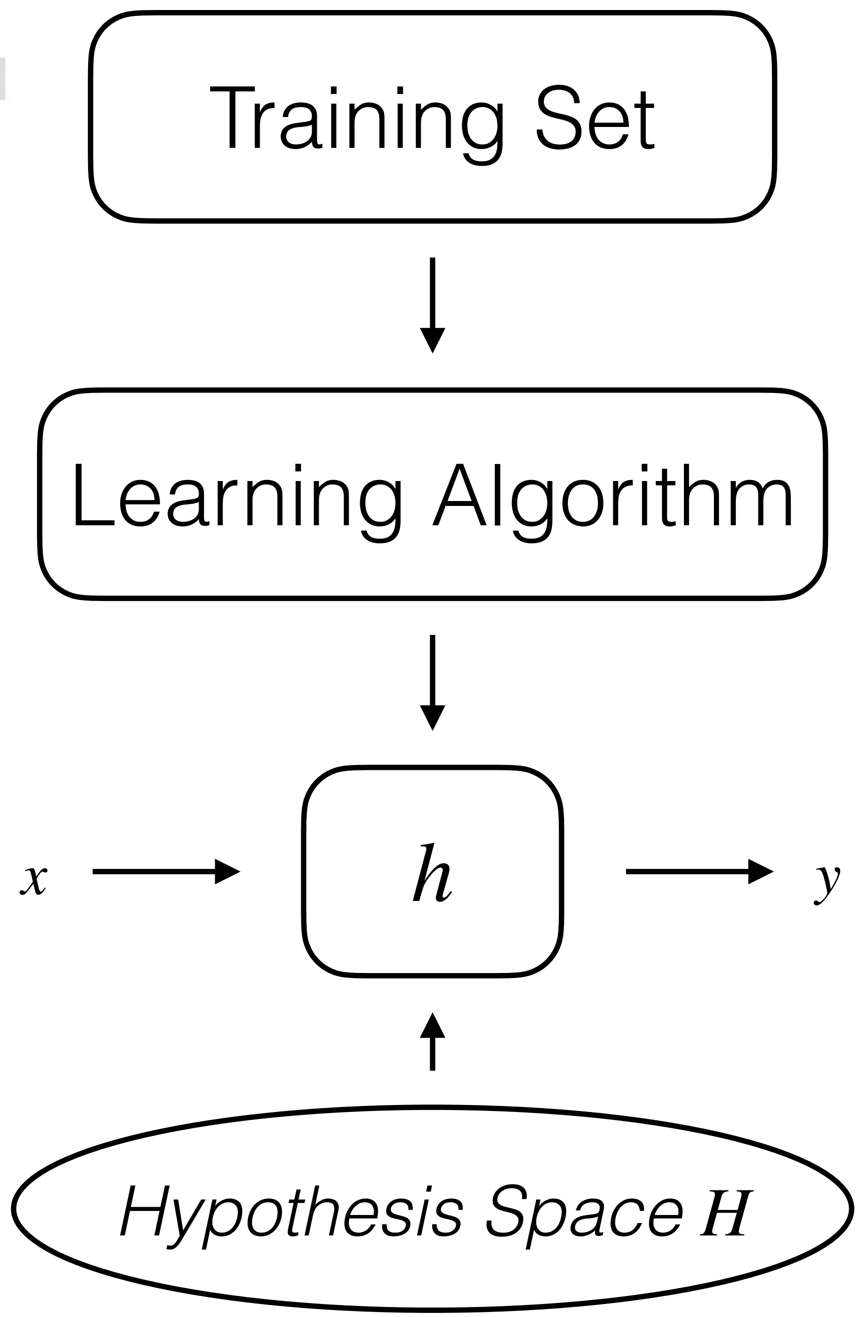
- **Supervised Learning**
  - **Goal:** give the "right answer" for each example in the data
  - Given examples $\{(x^{(i)}, y^{(i)})\}$, learn a function (description) which captures the information content of the examples
  - Basically we look for a function $h(\cdot)$ which is able to map in a predictive way $x^{(i)}$'s to $y^{(i)}$'s, *i.e.* $h: X \rightarrow Y$
  - An expert (or teacher) provides the supervision
    (i.e. the values of $h(\cdot)$ corresponding to the instances $x^{(i)}$)
  - **Output:** Classification (discrete-valued) *vs* Regression (real-valued output)

Main Learning Paradigms

- **Supervised Learning**

$\{(x^{(i)}, y^{(i)})\}$

Size in feet² (x)	Price ($) in 1K's (y)
2104	460
1416	232
1534	315
852	178
…	…

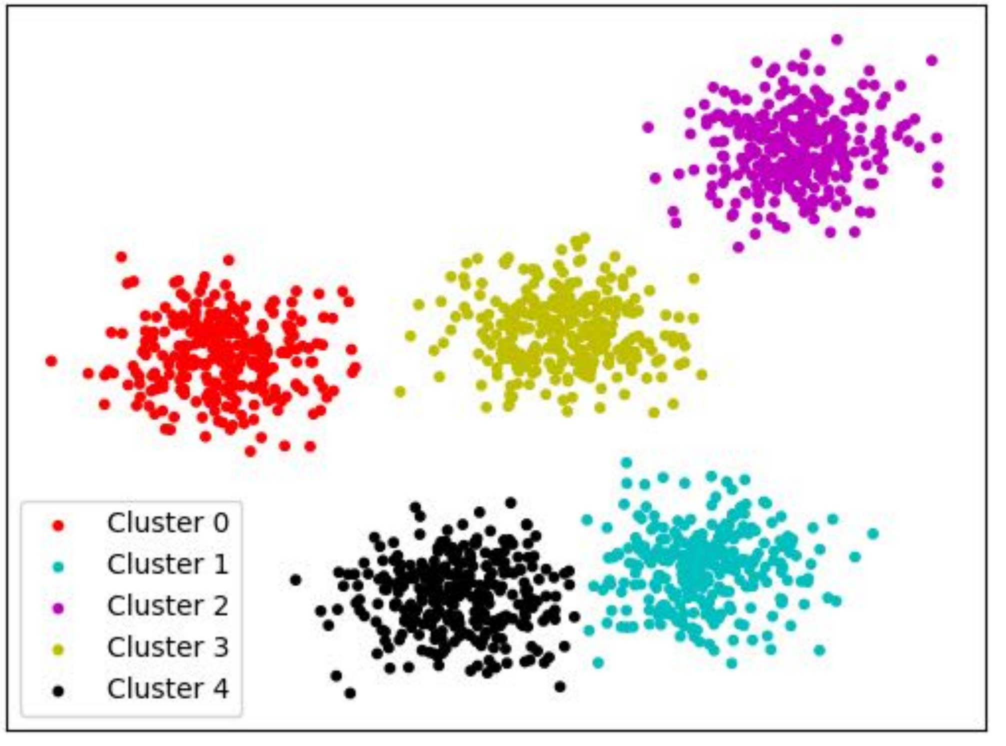


h approximates the unknown target f

$$h \sim f: X \rightarrow Y$$

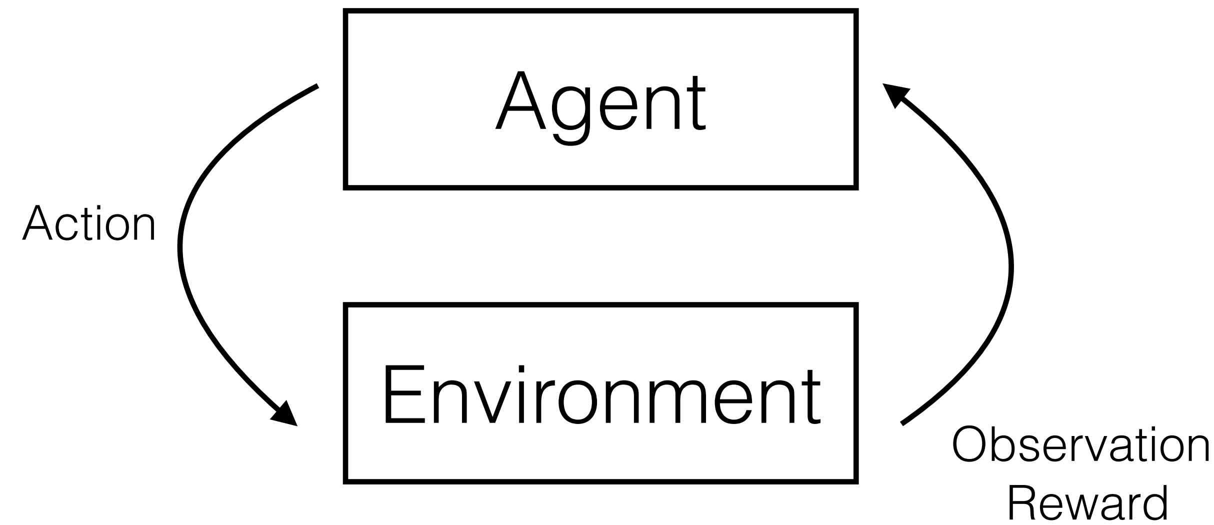
Main Learning Paradigms

- **Unsupervised Learning**
  - **Goal:** find regularities / patterns on the data
  - Given examples $\{x^{(i)}\}$, discover regularities on the whole input domain
  - There is no expert (*i.e.* no supervision)



Main Learning Paradigms

- **Reinforcement Learning**
  - Agent which may
    - be in state s
    - execute action a
      (among the ones admissible in state s)



  - and operates in an environment e, which in response to action a in the state s returns
    - the next state and a reward *r* (which can be positive, negative or neutral)
  - The goal of the agent is to maximize a function of the rewards

Other Learning Strategies

- Active Learning
- Online Learning and Incremental Learning
- Weak-supervised Learning
- Self-supervised Learning
- Deep Learning and Representation Learning
- Federated Learning

IML: Syllabus

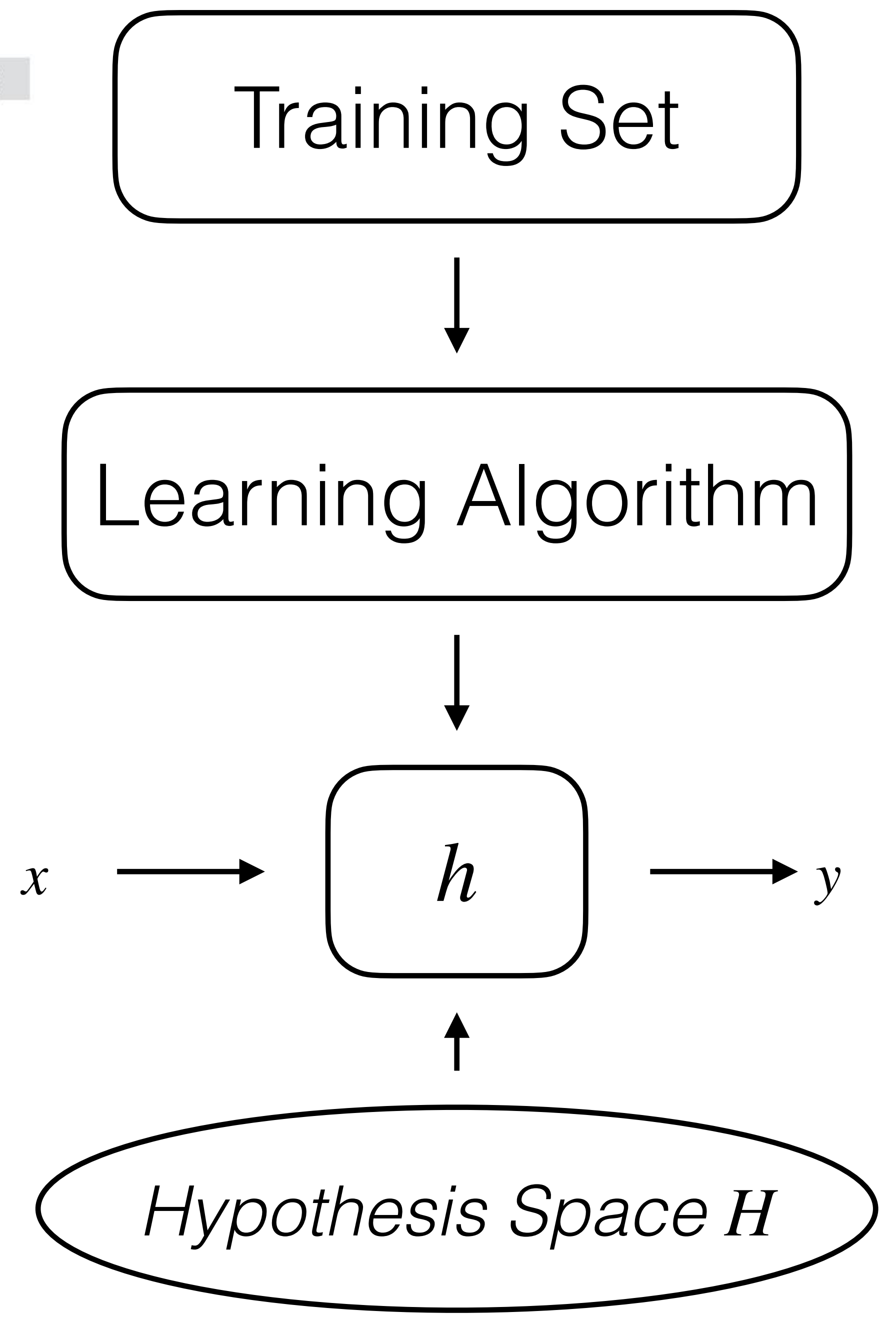
- Introduction: what is ML, main learning paradigms
- Linear regression, linear classification, logistic regression
- Neural Networks
- Evaluating a learning algorithm, model selection, boundaries
- Non-parametric methods, decision trees
- Probabilistic classifiers, Naive Bayes
- SVM, Kernel Methods
- Unsupervised Learning: clustering, dimensionality reduction
- Applications (in vision and NLP), cognitive services / ML APIs

Our main focus: Supervised Learning

- Classification (discrete) *vs* Regression (real-valued output)

$\{(x^{(i)}, y^{(i)})\}$

Size in feet^2 (x)	Price ($) in 1K's (y)
2104	460
1416	232
1534	315
852	178
…	…



h approximates the unknown target f

$$h \sim f: X \rightarrow Y$$

Fundamental Ingredients

- Training data D (drawn from the instance space X)
- Hypothesis space H
  - i.e. the set of functions which can be implemented by the machine learning system
  - we assume that the function to be learned f may be represented/approximated by the hypothesis $h \in H$
- Learning algorithm
  - It can be seen as a search algorithm into H

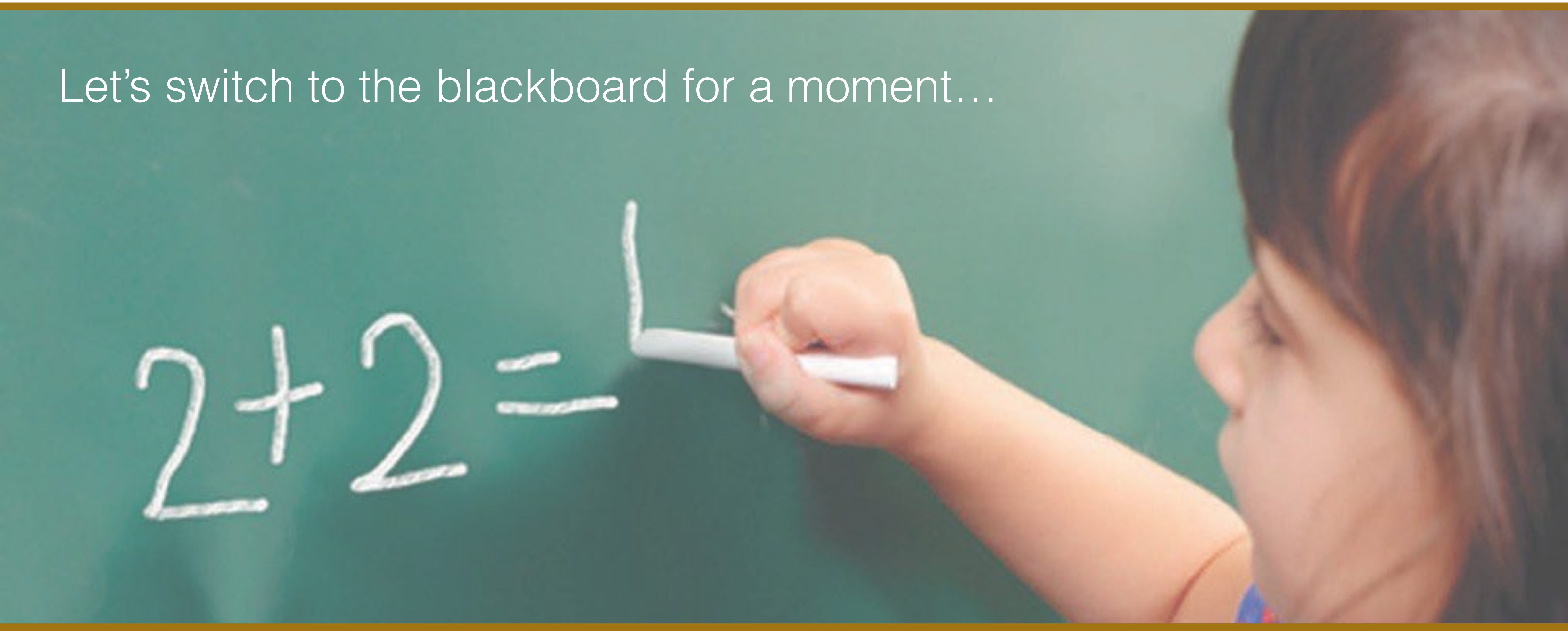
Inductive (or learning) Bias: on the representation (H) and/or on the search (learning algorithm)

Inductive Bias

- Experience alone might not allow us to make conclusions about unseen data instances
- **Inductive Bias** = all the assumptions about the "nature" of the target function and its selection
- Two type of bias:
  - Restriction: limit the hypothesis space
  - Preference: impose ordering on hypothesis space

Example of Inductive Bias

- 1. Linear regression, 2. Nearest neighbors



Example of Inductive Bias

- 1. Linear regression, 2. Nearest neighbors

Linear regression: assume that the output or dependent variable is related to independent variable linearly (in the weights).

Nearest neighbors: assume that most of the cases in a small neighborhood in feature space belong to the same class.

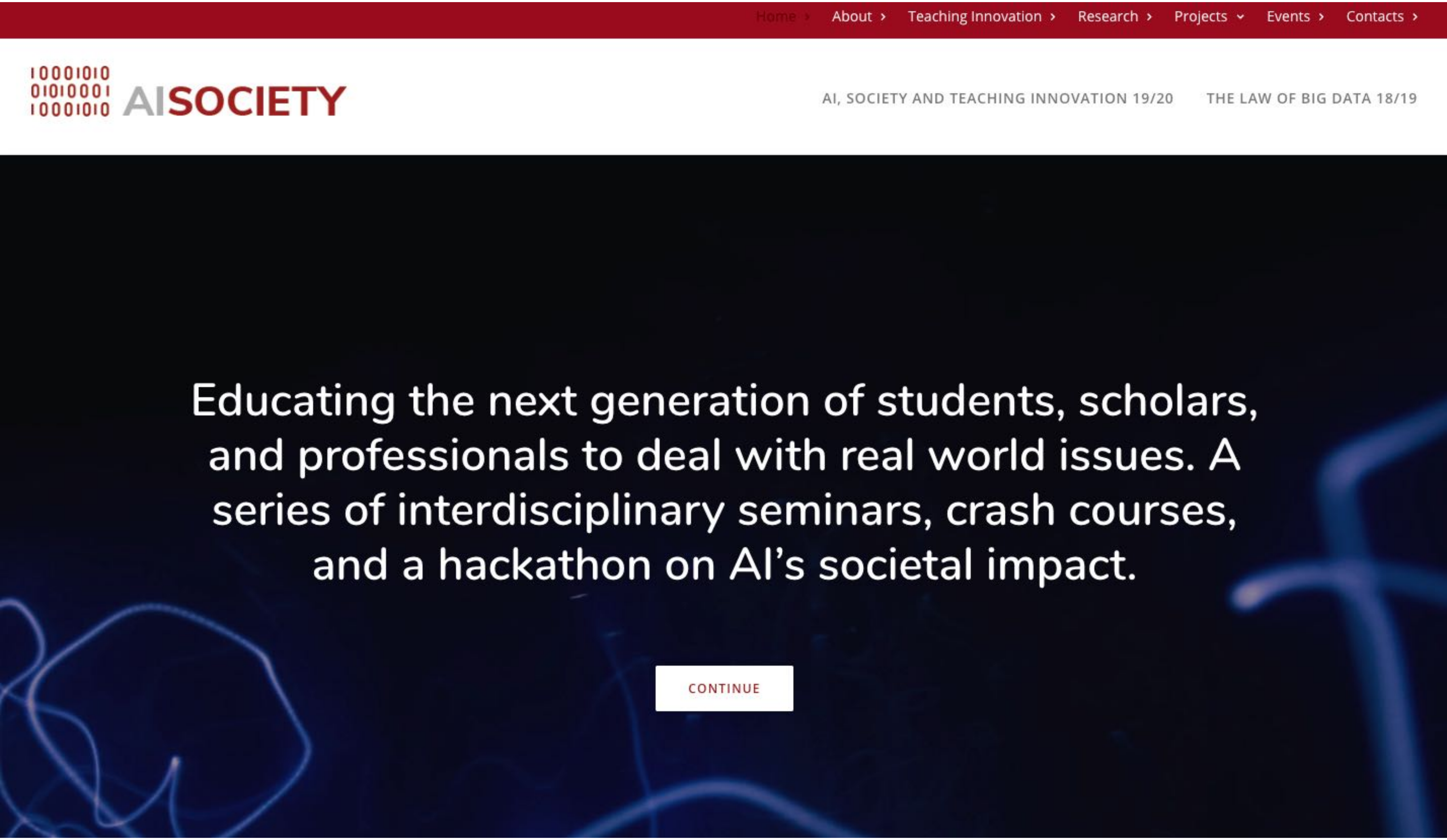
Algorithmic Bias

- It describes systematic and repeatable errors in a system that create unfair outcomes, such as privileging one arbitrary group of users over others
- This bias has been recently addressed also in legal frameworks (e.g. the 2018 European Union's GDPR)

Word embeddings (w2vNEWS)

Extreme *she*	Extreme *he*
1. homemaker	1. maestro
2. nurse	2. skipper
3. receptionist	3. protege
4. librarian	4. philosopher
5. socialite	5. captain
6. hairdresser	6. architect
7. nanny	7. financier
8. bookkeeper	8. warrior
9. stylist	9. broadcaster
10. housekeeper	10. magician

AI Society - seminar series & more



http://www.aisociety-unipd.it/

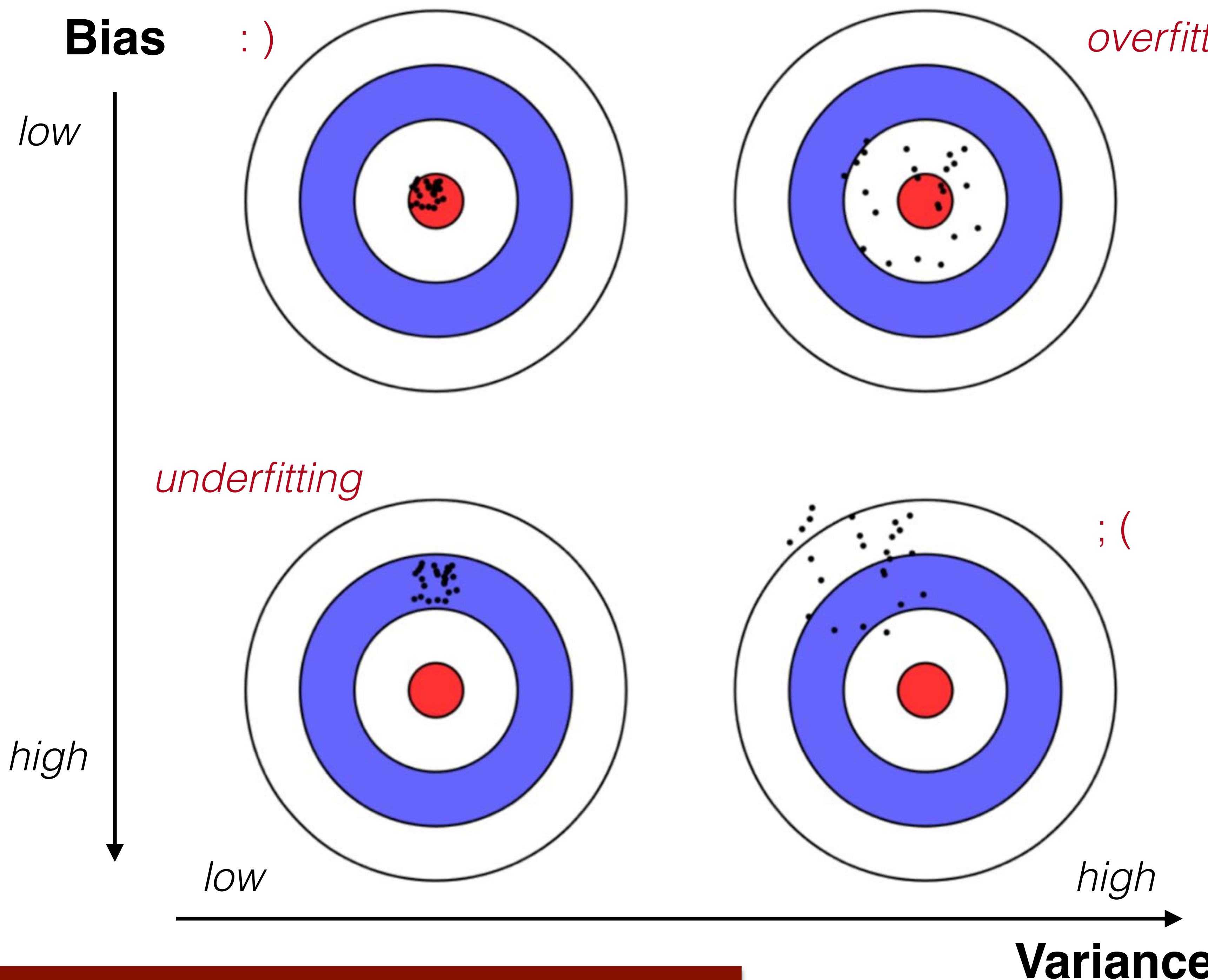
Bias-Variance Tradeoff

- The **bias** error is produced by weak assumptions in the learning algorithm
    - High bias can cause an algorithm to miss the relevant relations between features and target outputs (***underfitting***)

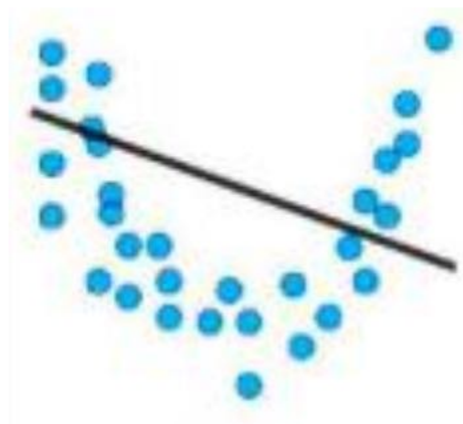
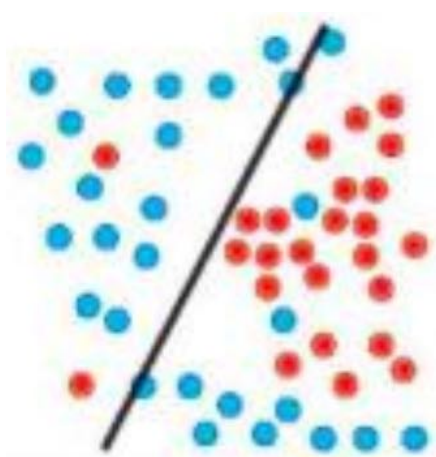
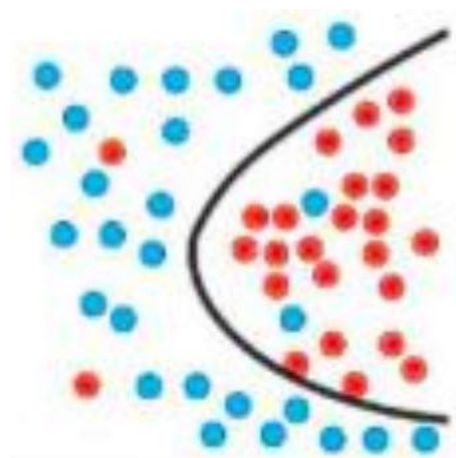
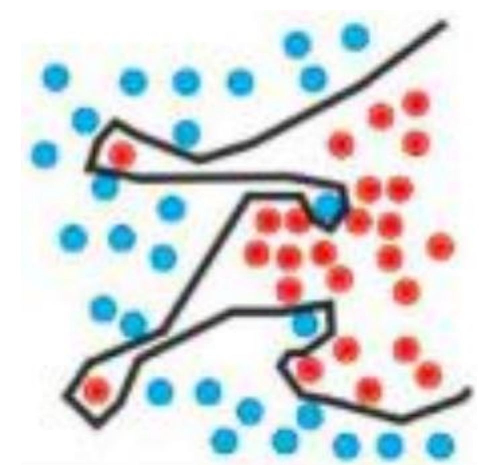
- The **variance** is an error produced by an over-sensitivity to small fluctuations in the training set
    - High variance can cause an algorithm to model the random noise in the training data, rather than the intended outputs (***overfitting***)

Bias-Variance Tradeoff

- Dartboard metaphor illustrating bias and variance:



Bias-Variance Tradeoff

	Underfitting	Optimal	Overfitting
Regression			
Classification			

Contact

- **Office:** Torre Archimede 6CD, room 622
- **Office hours** (ricevimento)**:** Friday 11:00-13:00

✉ lamberto.ballan@unipd.it

🏠 http://www.lambertoballan.net

🏠 http://vimp.math.unipd.it

@ twitter.com/lambertoballan