Introduction to Machine Learning

Fabio Massimo Zennaro

University of Bergen

This presentation will offer a **conceptual**/**critical** introduction to ML as possible **tool for research**:

An introduction to machine learning

This presentation will offer a **conceptual**/**critical** introduction to ML as possible **tool for research**:

✓ How to understand ML models;

An introduction to machine learning

This presentation will offer a **conceptual/critical** introduction to ML as possible **tool for research**:

- ✓ How to understand ML models;
- ✓ What are their strengths and limitations;

An introduction to machine learning

This presentation will offer a **conceptual**/**critical** introduction to ML as possible **tool for research**:

- ✓ How to understand ML models;
- ✓ What are their strengths and limitations;
- ✓ How to approach ML in research.

An introduction to machine learning

This presentation will offer a **conceptual**/**critical** introduction to ML as possible **tool for research**:

- ✓ How to understand ML models;
- ✓ What are their strengths and limitations;
- ✓ How to approach ML in research.

- × Hands-on tutorial;

An introduction to machine learning

This presentation will offer a **conceptual**/**critical** introduction to ML as possible **tool for research**:

- ✓ How to understand ML models;
- ✓ What are their strengths and limitations;
- ✓ How to approach ML in research.

- × Hands-on tutorial;
- × State-of-the-art review;

An introduction to machine learning

This presentation will offer a **conceptual**/**critical** introduction to ML as possible **tool for research**:

- ✓ How to understand ML models;
- ✓ What are their strengths and limitations;
- ✓ How to approach ML in research.

- × Hands-on tutorial;
- × State-of-the-art review;
- × LLMs.

A graphical roadmap

We will develop an understanding through a set of steps:

1. ML as an Oracle

$$X \longrightarrow Y$$

A graphical roadmap

We will develop an understanding through a set of steps:

1. *ML as an Oracle*

$$X \longrightarrow Y$$

2. *ML as a Black Box*

A graphical roadmap

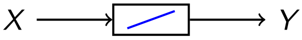
We will develop an understanding through a set of steps:

1. *ML as an Oracle*

$$X \longrightarrow Y$$

2. *ML as a Black Box*

3. *ML as an Induction Box*



A graphical roadmap

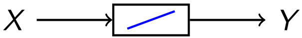
We will develop an understanding through a set of steps:

1. *ML as an Oracle*

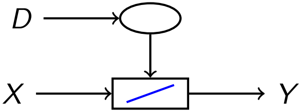
$$X \longrightarrow Y$$

2. *ML as a Black Box*

3. *ML as an Induction Box*



4. *ML as an Induction Process*



A graphical roadmap

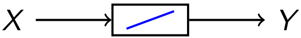
We will develop an understanding through a set of steps:

1. *ML as an Oracle*

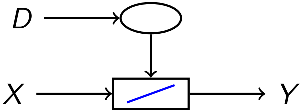
$$X \longrightarrow Y$$

2. *ML as a Black Box*

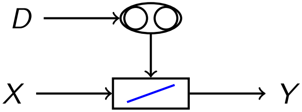
3. *ML as an Induction Box*



4. *ML as an Induction Process*



5. *ML as an Induction Algorithm*



1. ML as an Oracle

ML oracle

The most abstract *end-user* perspective on ML is:

$$\text{Data} \longrightarrow \text{Answers}$$

or even:

$$\text{Data} \longrightarrow \text{Knowledge}$$

Research and oracles

We do not want **oracles** in science:

Research and oracles

We do not want **oracles** in science:

- We are ignoring *complexities* and *subtleties* of ML.
  - We are remitting to an *authority principle*.

Research and oracles

We do not want **oracles** in science:

- We are ignoring *complexities* and *subtleties* of ML.
  - We are remitting to an *authority principle*.
- We just hope to get the right answers.
  - *Choosing which question* to ask is an important part of doing research.

2. ML as a Black Box

ML Black-Box

A first refinement comes in recognizing that there is something in between input and output:

This *black-box* relates (meaningfully) inputs and outputs.

ML Black-Box

A first refinement comes in recognizing that there is something in between input and output:

This *black-box* relates (meaningfully) inputs and outputs.

So far, it is a *magic box*.

ML Black-Box

How can inputs and outputs be connected?

ML Black-Box

How can inputs and outputs be connected?

The **mapping** relies on:

ML Black-Box

How can inputs and outputs be connected?

The **mapping** relies on:

- *Commonalities* in the input and output data;

ML Black-Box

How can inputs and outputs be connected?

The **mapping** relies on:

- *Commonalities* in the input and output data;
- *Common structures* in the input and output data;

ML Black-Box

How can inputs and outputs be connected?

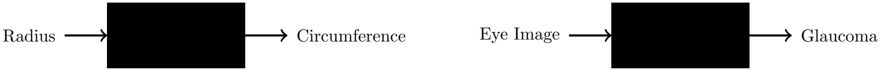
The **mapping** relies on:

- *Commonalities* in the input and output data;
- *Common structures* in the input and output data;
- *Correlation* in the input and output data;

Positive Examples

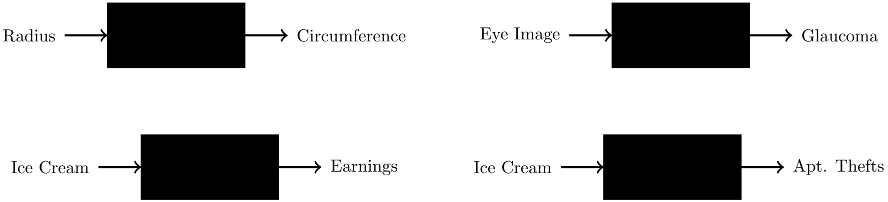
Positive Examples

Positive Examples

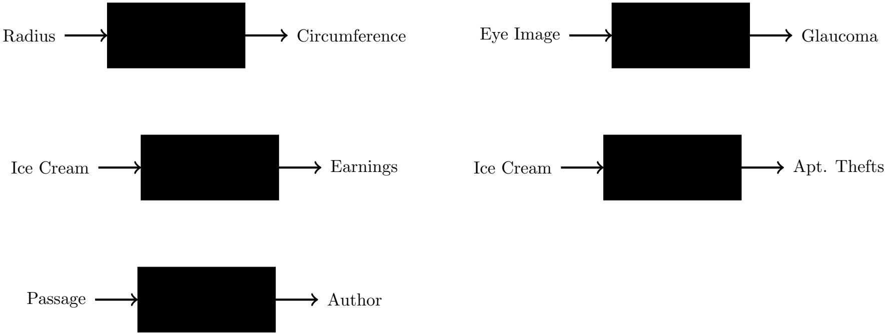


Positive Examples

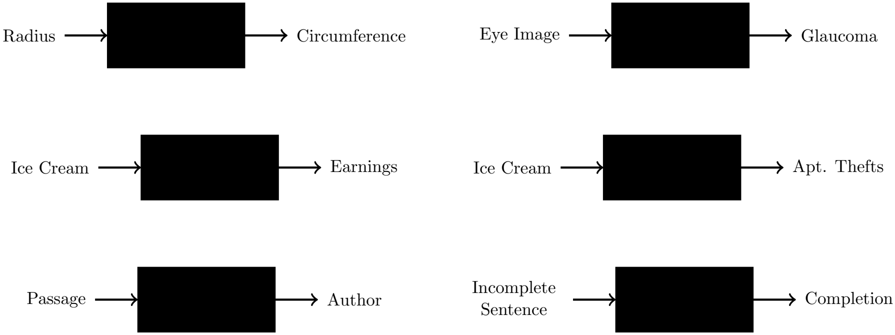
Positive Examples



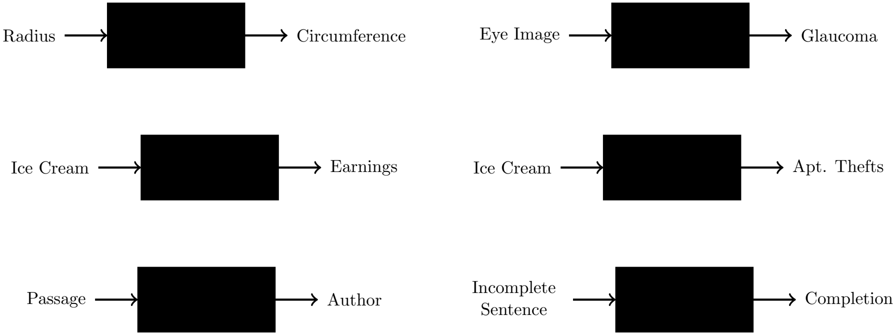
Positive Examples



Positive Examples



Positive Examples



All these *input-output pairings* were shown to work successfully;

i.e., there is a **correlation** between these inputs and outputs.

Negative Examples

Negative Examples

Negative Examples

This *input-output pairing* has no correlation - as far as we know.

It does not work.

Negative Examples

This *input-output pairing* has no correlation - as far as we know.

It does not work.

Negative Examples

This *input-output pairing* has no correlation - as far as we know.

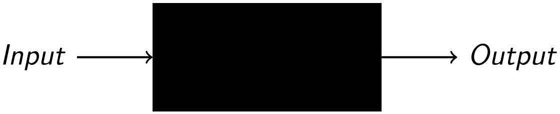
It does not work.

This *input-output pairing* has correlation to a certain degree.

Should we use it?

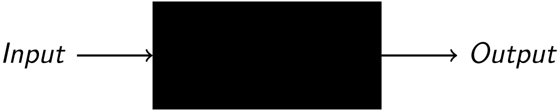
Research and black-boxes

This simple understanding already tells us something about the *possibilities* of ML.



Research and black-boxes

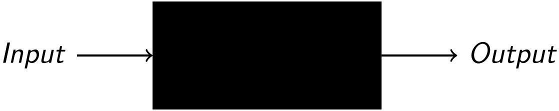
This simple understanding already tells us something about the *possibilities* of ML.



It allows us to evaluate when:

Research and black-boxes

This simple understanding already tells us something about the *possibilities* of ML.

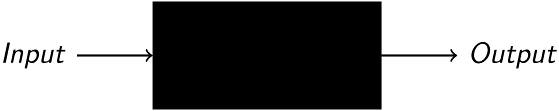


It allows us to evaluate when:

- we can expect ML *to work*;

Research and black-boxes

This simple understanding already tells us something about the *possibilities* of ML.

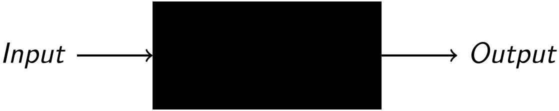


It allows us to evaluate when:

- we can expect ML *to work*;
- we can expect ML *not to work*;

Research and black-boxes

This simple understanding already tells us something about the *possibilities* of ML.



It allows us to evaluate when:

- we can expect ML *to work*;
- we can expect ML *not to work*;
- we may *not want to use* ML.

3. ML as an Induction Box

ML Method

The black-box perspective, still, is very superficial.

ML Method

The black-box perspective, still, is very superficial.

We answered *what* ML does (connect input and outputs through their patterns), but not *how*.

ML Method

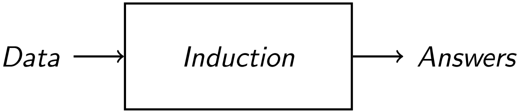
The black-box perspective, still, is very superficial.

We answered *what* ML does (connect input and outputs through their patterns), but not *how*.

We still want to understand the **method** hidden behind the black box.

ML Induction Machine

The ML box can be thought of as an **induction machine**.



ML Induction Machine

The ML box can be thought of as an **induction machine**.

$$\textit{Data} \longrightarrow \boxed{\textit{Induction}} \longrightarrow \textit{Answers}$$

From limited *observations* it *induces* a *general principle*.

Induction and Generalization

A critical aspect of *induction* is **generalization**:

Induction and Generalization

A critical aspect of *induction* is **generalization**:

- We do not want just to *memorize* observations;

Induction and Generalization

A critical aspect of *induction* is **generalization**:

- We do not want just to *memorize* observations;
- We want to **generalize** to unseen events.

Induction and Generalization

A critical aspect of *induction* is **generalization**:

- We do not want just to *memorize* observations;
- We want to **generalize** to unseen events.

As a human being, from the sun rising every day, I predict it will rise tomorrow too.

Induction and Generalization

A critical aspect of *induction* is **generalization**:

- We do not want just to *memorize* observations;
- We want to **generalize** to unseen events.

As a human being, from the sun rising every day, I predict it will rise tomorrow too.

As a doctor, from observing many cases of glaucoma, I learn to distinguish new instances.

Problem of Induction

What guarantees do we have that *generalization* is true?

> *"It implies no contradiction that the course of nature may change, and that an object seemingly like those which we have experienced, may be attended with different or contrary effects." (David Hume)*[1]

[1] D. Hume, *An Enquiry Concerning Human Understanding*

Problem of Induction

What guarantees do we have that *generalization* is true?

> *"It implies no contradiction that the course of nature may change, and that an object seemingly like those which we have experienced, may be attended with different or contrary effects." (David Hume)*[1]

We need to assume a **principle of uniformity of nature**.

[1] D. Hume, *An Enquiry Concerning Human Understanding*

Limits of the Induction

What are the **limits** of the *principle of uniformity*?

Limits of the Induction

What are the **limits** of the *principle of uniformity*?

I can say that the sun will rise tomorrow, in a month, and in a year. But in five billion years?

Limits of the Induction

What are the **limits** of the *principle of uniformity*?

I can say that the sun will rise tomorrow, in a month, and in a year. But in five billion years?

As a doctor, at what point should an unusual image stop me from drawing conclusions from experience?

Induction for ML

To understand how ML performs inductions we need to understand:

(A) How an *inductive model* is implemented?
(B) What role is the *principle of uniformity* playing?
(C) How *generalization* arises?
(D) What are the *limits of induction* for ML?

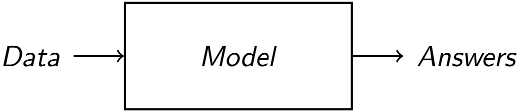
(A) Induction for ML

Practically, ML performs induction over data by generating a **model** of the data.

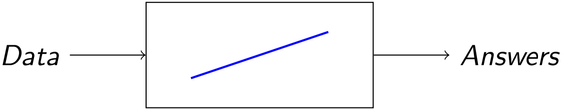
$$\textit{Data} \longrightarrow \boxed{\quad \textit{Model} \quad} \longrightarrow \textit{Answers}$$

(A) Induction for ML

Practically, ML performs induction over data by generating a **model** of the data.



Substantially, such a model is a mathematical **function**:



(Reminder: Function)

A *mathematical function* is a law mapping an input to a single output:

(Reminder: Function)

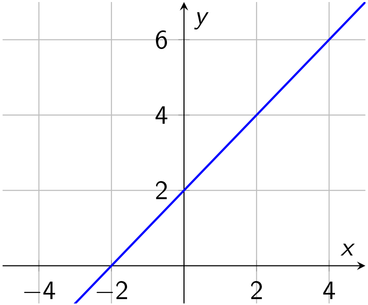
A *mathematical function* is a law mapping an input to a single output:

$$f : X \to Y$$

(Reminder: Function)

A *mathematical function* is a law mapping an input to a single output:

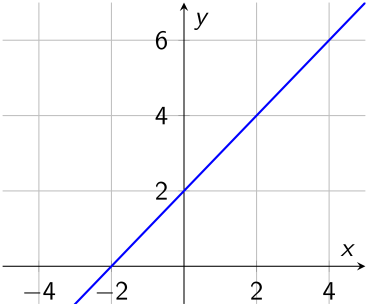
$$f : X \to Y$$



(We draw a linear function for simplicity. It does not have to be linear.)

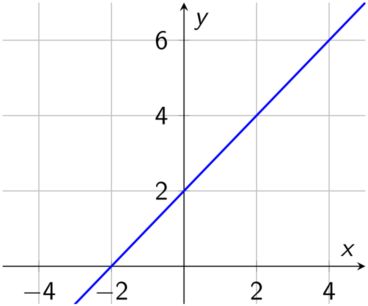
(B) Functionality Assumption

ML claims that that there exists a **true function** or **true generating process** f^* that generated the data:



(B) Functionality Assumption

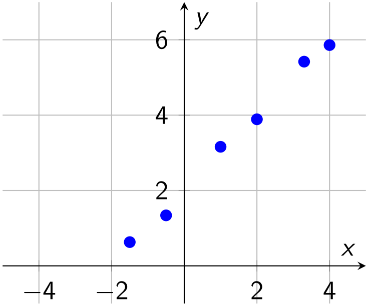
ML claims that that there exists a **true function** or **true generating process** f^* that generated the data:



This claim is the realization of the **principle of uniformity of nature**!

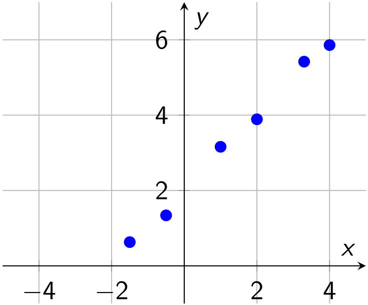
(C) Functionality Assumption

In reality, ML is only given *datapoints* from the true function f^*:



(C) Functionality Assumption

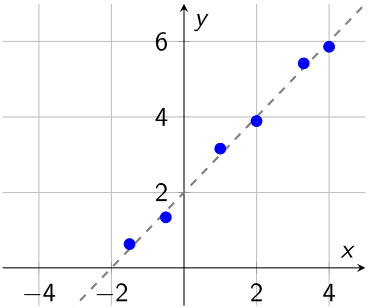
In reality, ML is only given *datapoints* from the true function f^*:



These are our **observations**.

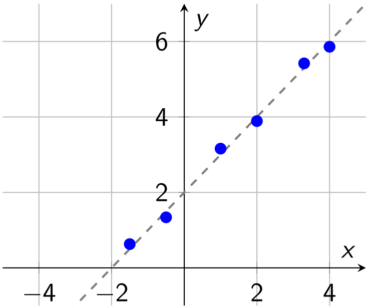
(C) Functionality Assumption

ML claims it will *recover* a function $\hat{f}$ close to the true function f^*:



(C) Functionality Assumption

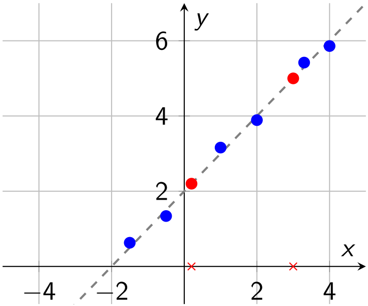
ML claims it will *recover* a function $\hat{f}$ close to the true function f^*:



This is the actual **induction**.

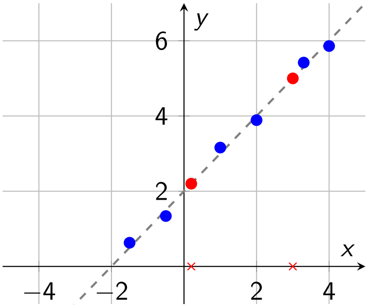
(C) Generalization

Consequently, ML claims its model *predicts* all the points provided by the true f^*:



(C) Generalization

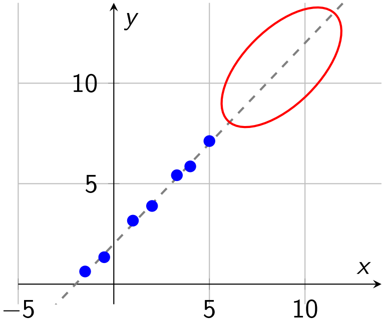
Consequently, ML claims its model *predicts* all the points provided by the true f^*:



This is the actual **generalization**.

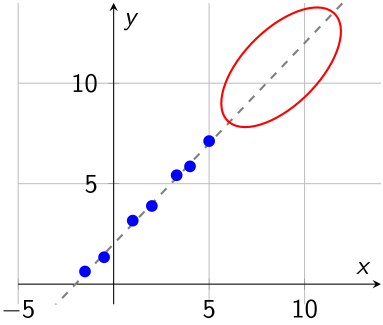
(D) Domain (of Uniformity)

Up to where does the generalization of the model hold?



(D) Domain (of Uniformity)

Up to where does the generalization of the model hold?

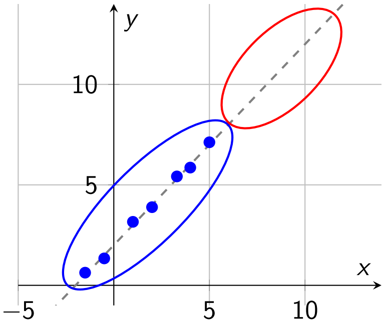


We want to identify the **limits** of generalization.

(D) Domain (of Uniformity)

We distinguish between a domain of:

- **Interpolation**: area covered by sample and induction;
- **Extrapolation**: area not covered by samples where induction is uncertain.

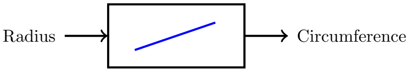


Positive Examples

Behind all positive examples, we postulate a *true generating process*.

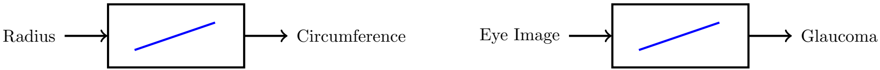
Positive Examples

Behind all positive examples, we postulate a *true generating process*.



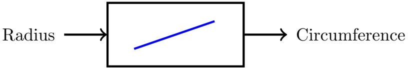
Positive Examples

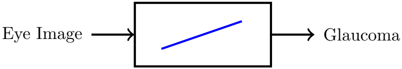
Behind all positive examples, we postulate a *true generating process*.

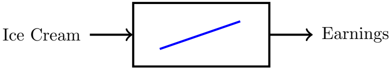


Positive Examples

Behind all positive examples, we postulate a *true generating process*.

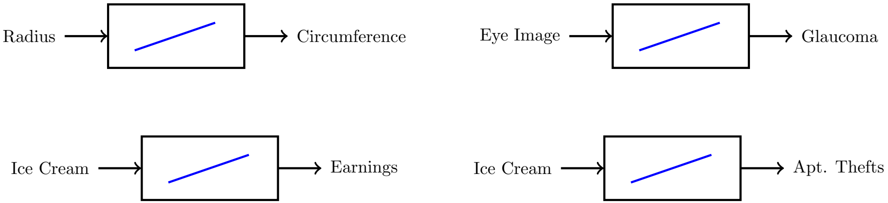






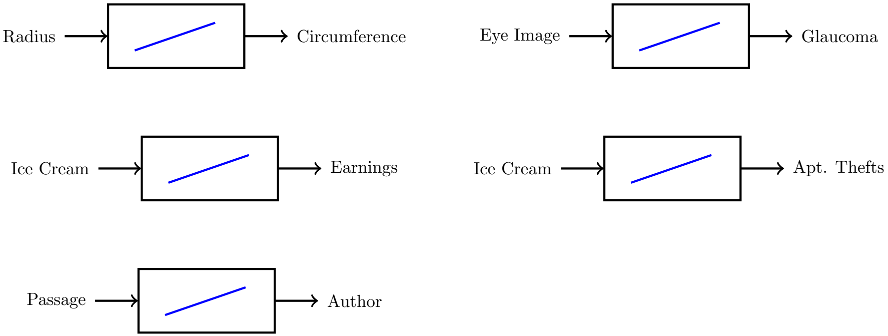
Positive Examples

Behind all positive examples, we postulate a *true generating process*.



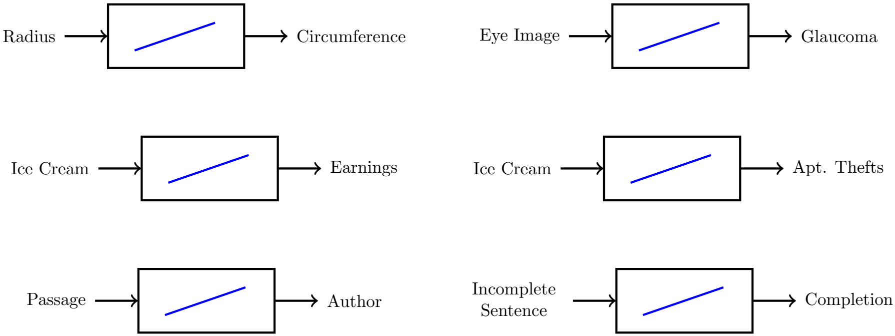
Positive Examples

Behind all positive examples, we postulate a *true generating process*.



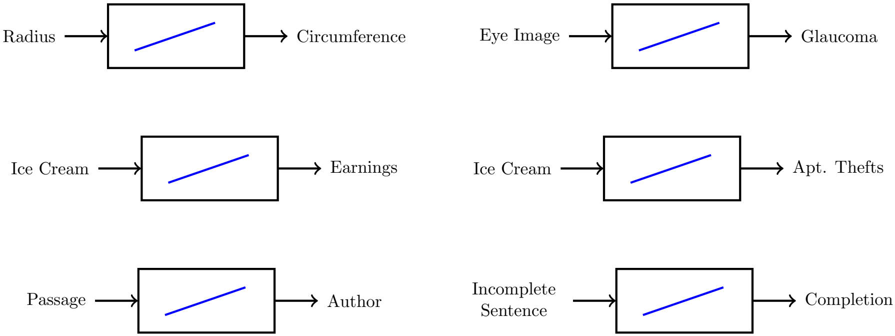
Positive Examples

Behind all positive examples, we postulate a *true generating process*.



Positive Examples

Behind all positive examples, we postulate a *true generating process*.



Given enough data, ML claims it will recover this *function*.

Research and induction

Induction is not the only way of reasoning, but it is **very important**.

[2] nature.com/articles/d41586-024-02441-2

Research and induction

Induction is not the only way of reasoning, but it is **very important**.

When we do **research** we rely greatly on induction:

[2] nature.com/articles/d41586-024-02441-2

Research and induction

Induction is not the only way of reasoning, but it is **very important**.

When we do **research** we rely greatly on induction:

- *Empirical sciences* know the world through senses, through data points

[2] nature.com/articles/d41586-024-02441-2

Research and induction

Induction is not the only way of reasoning, but it is **very important**.

When we do **research** we rely greatly on induction:

- *Empirical sciences* know the world through senses, through data points
- *Deductive sciences* often rely on induction in their steps[2]

[2] nature.com/articles/d41586-024-02441-2

Research and induction

Induction is not the only way of reasoning, but it is **very important**.

When we do **research** we rely greatly on induction:

- *Empirical sciences* know the world through senses, through data points
- *Deductive sciences* often rely on induction in their steps[2]

Reasoning does not (probably) reduce to induction, but induction is very important.

[2] `nature.com/articles/d41586-024-02441-2`

Research and induction

"All the impressive achievements of deep learning amount to just curve fitting." (Judea Pearl)

Research and induction

> *"All the impressive achievements of deep learning amount to just curve fitting." (Judea Pearl)*

Lots of debate about this characterization of ML: all this success and ML is *only curve fitting*?

Research and induction

> *"All the impressive achievements of deep learning amount to just curve fitting." (Judea Pearl)*

Lots of debate about this characterization of ML: all this success and ML is *only curve fitting*?

Yet the opposite perspective might be more interesting: simple curve fitting performs *amazing inductions*!

Research and induction

Using ML amounts to performing inductions, so it is worth considering:

Research and induction

Using ML amounts to performing inductions, so it is worth considering:

- *Is induction possible* between inputs and outputs of interest?
    - Is there a correlation between them?

Research and induction

Using ML amounts to performing inductions, so it is worth considering:

- *Is induction possible* between inputs and outputs of interest?
  - Is there a correlation between them?

- *What observations* would make such an induction possible?
  - What data should I collect?
  - Remember data do not have to be limited to human sensory modalities!

Research and induction

Using ML amounts to performing inductions, so it is worth considering:

- *Is induction possible* between inputs and outputs of interest?
  - Is there a correlation between them?
- *What observations* would make such an induction possible?
  - What data should I collect?
  - Remember data do not have to be limited to human sensory modalities!
- *How strong is the relation* between inputs and outputs?
  - How much data should I collect?

(Aside: Statistics)

Statistics has been concerned with induction for a long time.

(Aside: Statistics)

Statistics has been concerned with induction for a long time.

ML borrows heavily from statistics and integrates it with:

- *Optimization*
- *Operational Research*
- *Software Engineering*
- *Numerical Methods*
- ...

(Aside: Statistics)

Statistics has been concerned with induction for a long time.

ML borrows heavily from statistics and integrates it with:

- *Optimization*
- *Operational Research*
- *Software Engineering*
- *Numerical Methods*
- ...

This gives ML a very *applied* flavour.

(Aside: Statistics vs ML)

A *technical distinction* from *Geoffrey Hinton*[3]:

[3] G. Hinton, *Basic Machine Learning*

(Aside: Statistics vs ML)

A *technical distinction* from *Geoffrey Hinton*[3]:

A spectrum of machine learning tasks

Statistics-----------------------Artificial Intelligence

Statistics	Artificial Intelligence
• Low-dimensional data (e.g. less than 100 dimensions)	• High-dimensional data (e.g. more than 100 dimensions)
• Lots of noise in the data	• The noise is not sufficient to obscure the structure in the data if we process it right.
• There is not much structure in the data, and what structure there is, can be represented by a fairly simple model.	• There is a huge amount of structure in the data, but the structure is too complicated to be represented by a simple model.
• The main problem is distinguishing true structure from noise.	• The main problem is figuring out a way to represent the complicated structure that allows it to be learned.

[3] G. Hinton, *Basic Machine Learning*

(Aside: Statistics vs ML)

A *conceptual distinction* from *Jun Otsuka*[4]:

[4] J. Otsuka, *Thinking About Statistics*

(Aside: Statistics vs ML)

A *conceptual distinction* from *Jun Otsuka*[4]:

- **Statistics** is characterized by a principle of *truthfulness*.
    - We want models that *explain* reality.

[4] J. Otsuka, *Thinking About Statistics*

(Aside: Statistics vs ML)

A *conceptual distinction* from *Jun Otsuka*[4]:

- **Statistics** is characterized by a principle of *truthfulness*.
  - We want models that *explain* reality.
- **ML** is characterized by a principle of *utility*.
  - We want models that *predict* reality.

[4] J. Otsuka, *Thinking About Statistics*

(Aside: Statistics vs ML)

Both perspectives tell us something important if we want to use ML in research.

(Aside: Statistics vs ML)

Both perspectives tell us something important if we want to use ML in research.

- Hinton tells us that the *sort of data* we have may affect our use of ML;

(Aside: Statistics vs ML)

Both perspectives tell us something important if we want to use ML in research.

- Hinton tells us that the *sort of data* we have may affect our use of ML;
- Otsuka tells us that the *sort of knowledge* we want to achieve may justify our choice of ML.

(Aside: Why Intelligence?)

Why do we talk about *intelligence* in relation to this process of induction?

[5] M. Hutter, *Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability*

(Aside: Why Intelligence?)

Why do we talk about *intelligence* in relation to this process of induction?

We can see *models* as *summaries*.

[5] M. Hutter, *Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability*

(Aside: Why Intelligence?)

Why do we talk about *intelligence* in relation to this process of induction?

We can see *models* as *summaries*.

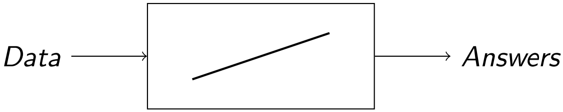
Compression is (part of) intelligence[5].

[5] M. Hutter, *Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability*

4. ML as an Induction Process

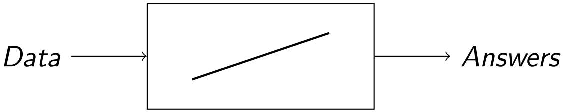
Static ML

The current understanding of ML has still one limitation:



Static ML

The current understanding of ML has still one limitation:

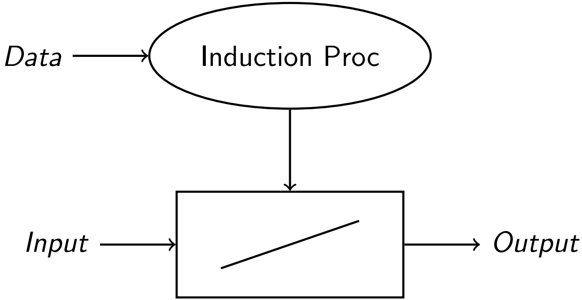


We see ML as a **static** box containing a function performing a successful induction.

How do we get to that function?

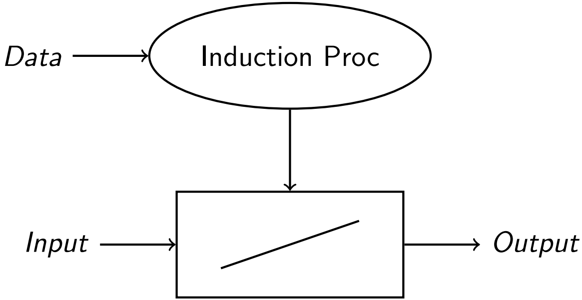
ML Induction Process

ML implies a **dynamic** process where data points are progressively observed and the function $\hat{f}$ is refined:



ML Induction Process

ML implies a **dynamic** process where data points are progressively observed and the function $\hat{f}$ is refined:



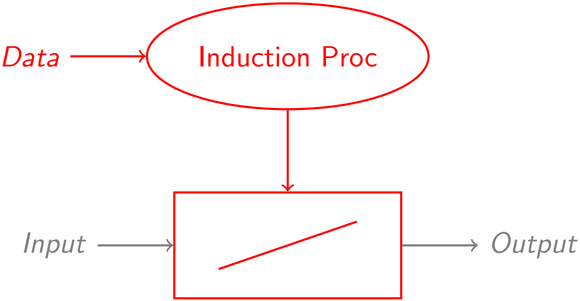
We have two main phases in this dynamics:

(A) *training*/*learning phase*;

(B) *prediction*/*inference phase*.

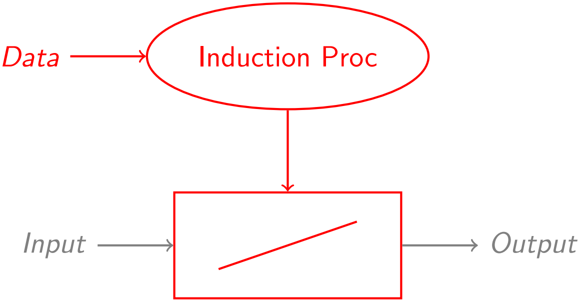
(A) Learning Phase

In **learning phase** we process the data to produce a model.



(A) Learning Phase

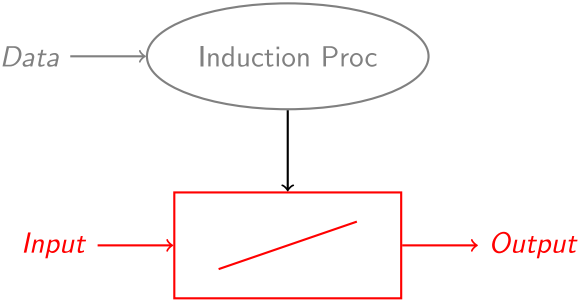
In **learning phase** we process the data to produce a model.



This is the real challenging phase.

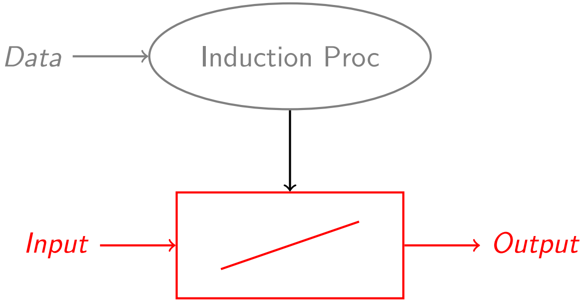
(B) Prediction Phase

In **prediction phase** we use the model to draw conclusions.



(B) Prediction Phase

In **prediction phase** we use the model to draw conclusions.



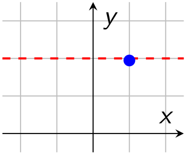
This is the use phase.

Learning Process

In the learning phase, the ML algorithm tries to progressively *fit the observations*.

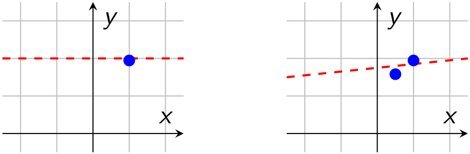
Learning Process

In the learning phase, the ML algorithm tries to progressively *fit the observations*.



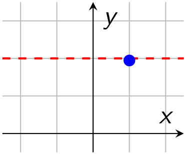
Learning Process

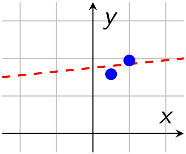
In the learning phase, the ML algorithm tries to progressively *fit the observations*.



Learning Process

In the learning phase, the ML algorithm tries to progressively *fit the observations*.

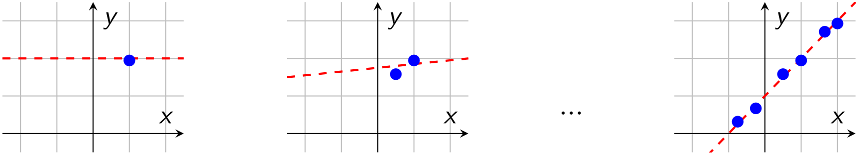




...

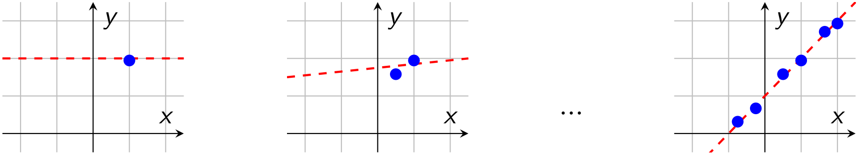
Learning Process

In the learning phase, the ML algorithm tries to progressively *fit the observations*.



Learning Process

In the learning phase, the ML algorithm tries to progressively *fit the observations*.



The ML algorithm is the **induction engine**.

Research and induction process

Seeing induction as a process invites some considerations:

Research and induction process

Seeing induction as a process invites some considerations:

- *What data* am I using to drive the induction?
  - Are the data consistent?
  - (Are the data i.d., from the same distribution or process?)

Research and induction process

Seeing induction as a process invites some considerations:

- *What data* am I using to drive the induction?
  - Are the data consistent?
  - (Are the data i.d., from the same distribution or process?)

- Am I providing *data covering my domain of interest*?

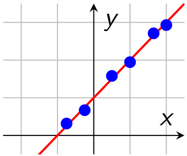
5. ML as an Induction Algorithm

Learning Process

Discussing about a learning process raises many questions:

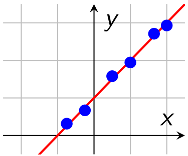
Learning Process

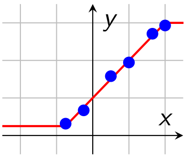
Discussing about a learning process raises many questions:



Learning Process

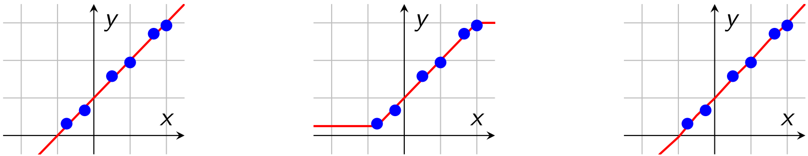
Discussing about a learning process raises many questions:





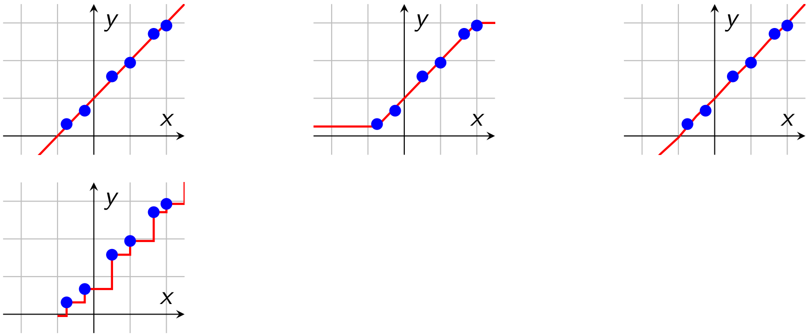
Learning Process

Discussing about a learning process raises many questions:



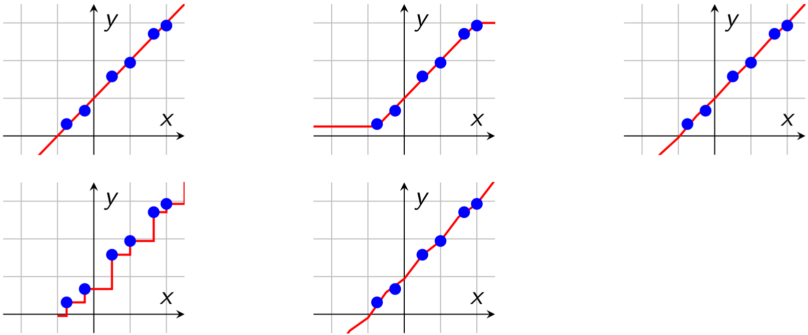
Learning Process

Discussing about a learning process raises many questions:



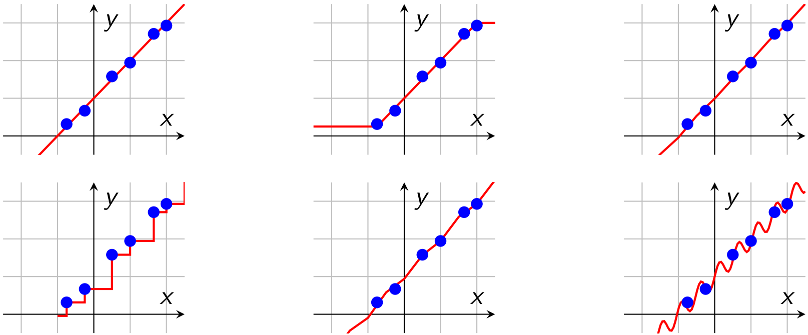
Learning Process

Discussing about a learning process raises many questions:



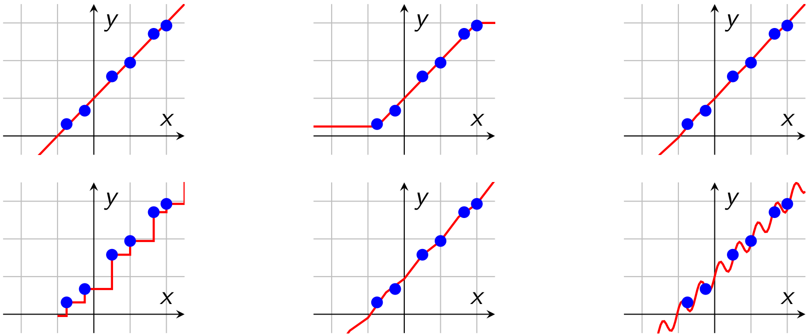
Learning Process

Discussing about a learning process raises many questions:



Learning Process

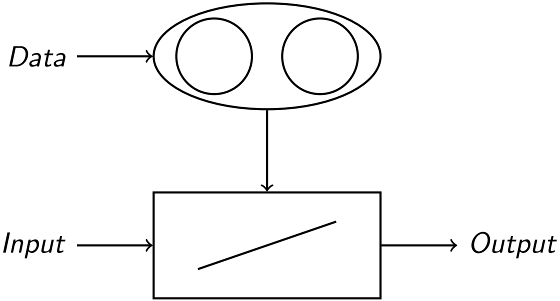
Discussing about a learning process raises many questions:



- *What is the right fitting?*
- *When do we know we have reached the correct fitting?*

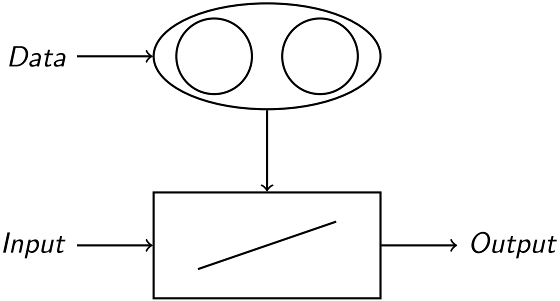
Anatomy of ML Algorithm

We want to look into the *induction process*:



Anatomy of ML Algorithm

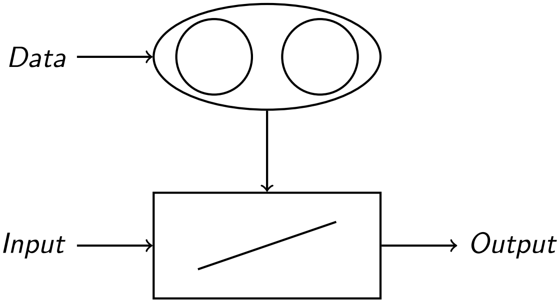
We want to look into the *induction process*:



We want to get a grasp on the **algorithm** driving the induction process.

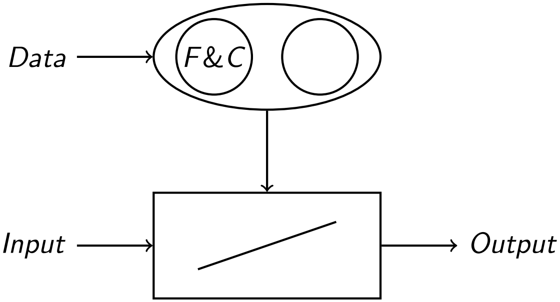
Anatomy of ML Algorithm

There are a couple of *components* of a ML algorithm that determine the result of induction:



Anatomy of ML Algorithm

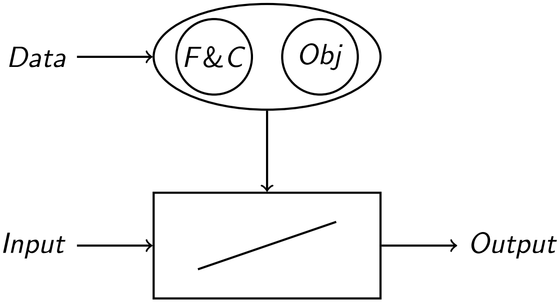
There are a couple of *components* of a ML algorithm that determine the result of induction:



(A) **Family of models and constraints**

Anatomy of ML Algorithm

There are a couple of *components* of a ML algorithm that determine the result of induction:



(A) **Family of models and constraints**

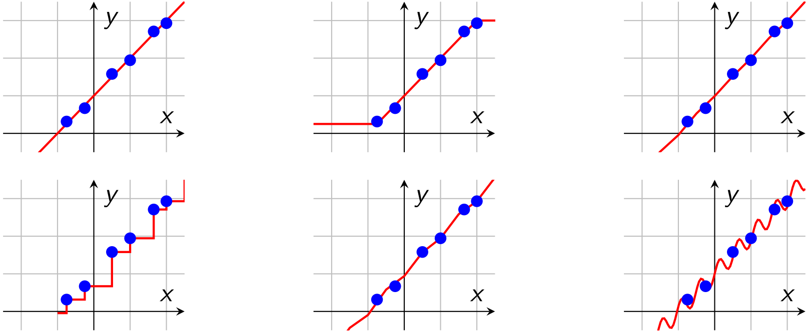
(B) **Objective/Cost**

(A) Learning Process

We can not decide, based on the data, which is the *true generative process*...

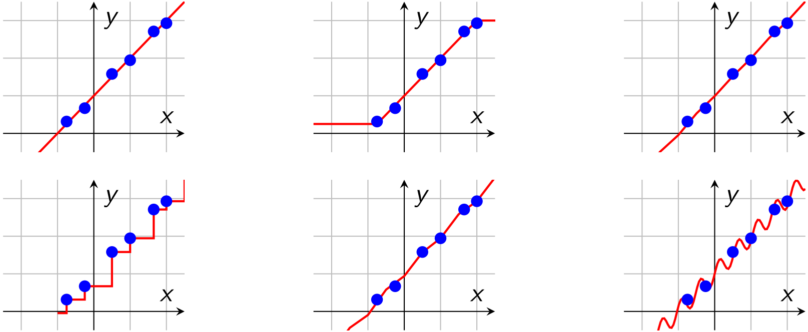
(A) Learning Process

We can not decide, based on the data, which is the *true generative process*...



(A) Learning Process

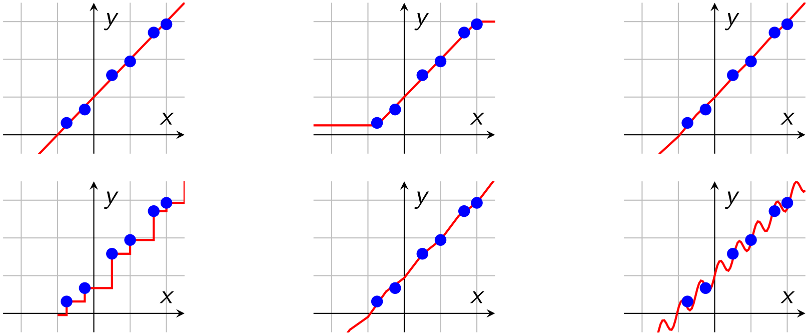
We can not decide, based on the data, which is the *true generative process*...



There is a **huge space** of possible models!

(A) Learning Process

We can not decide, based on the data, which is the *true generative process*...



There is a **huge space** of possible models!

We need to make some *assumptions*.

(A) Simplicity Assumption

> *"Entities are not to be multiplied without necessity." (William of Ockham/John Punch)*

The *simplest* explanation is the *most likely* one.

(A) Simplicity Assumption

> *"Entities are not to be multiplied without necessity." (William of Ockham/John Punch)*

The *simplest* explanation is the *most likely* one.

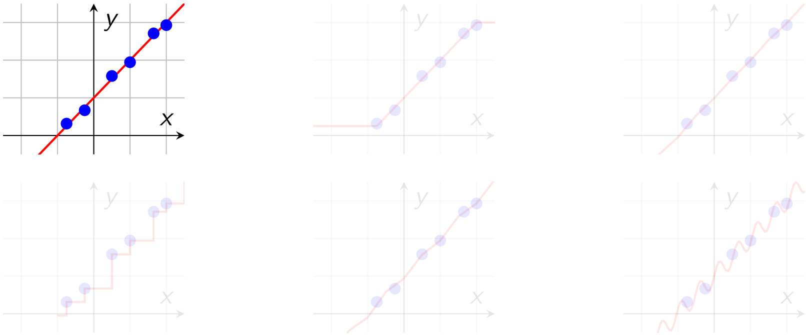
Overly complex models risk postulating non-existent dynamics.

(A) Simple Models

For instance, **linear models** are very simple:

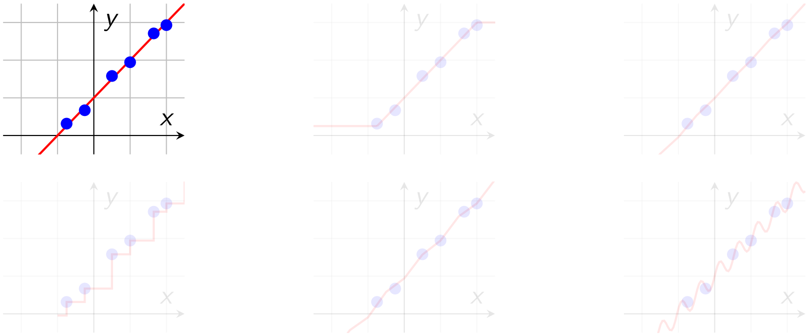
(A) Simple Models

For instance, **linear models** are very simple:



(A) Simple Models

For instance, **linear models** are very simple:



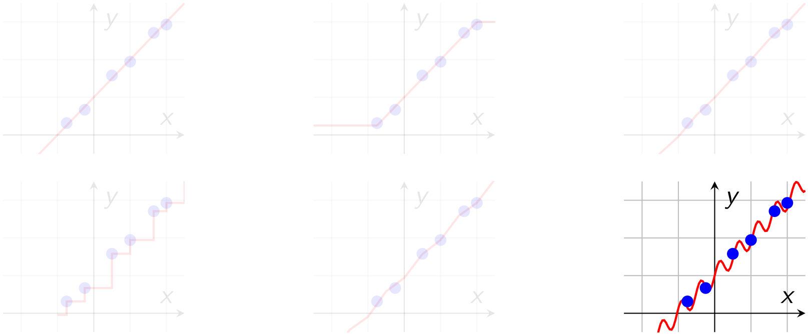
Alternatively, we could consider models that are *quadratic, polynomial, piecewise, sinusoidal...*

(A) Complex Models

Complex models may capture the datapoints as well as noise:

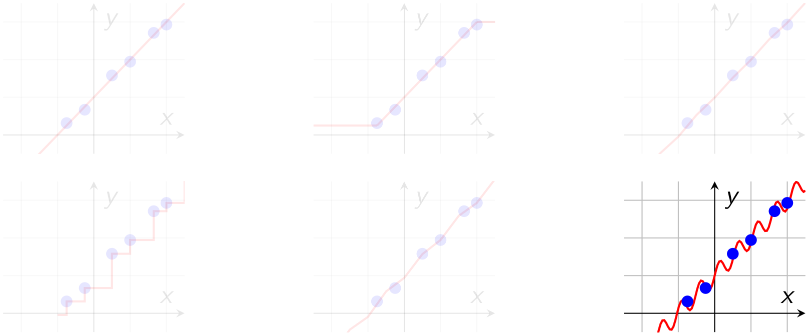
(A) Complex Models

Complex models may capture the datapoints as well as noise:



(A) Complex Models

Complex models may capture the datapoints as well as noise:



This is also called **overfitting**.

(A) Family of models

By restricting the **family of models** we:

- impose restrictions on *complexity*;
- restrict the *solution space*.

(A) Flexible Models

An alternative approach is to rely on **flexible models**.

(A) Flexible Models

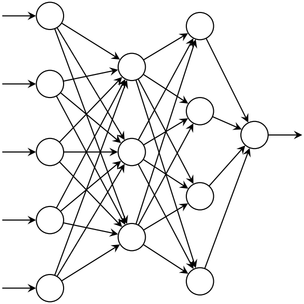
An alternative approach is to rely on **flexible models**.

Let the model encompass many families and adapt to the data.

(Implicitly, *high complexity*)

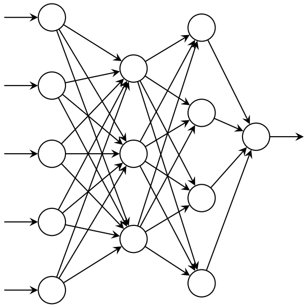
(A) Neural Networks

A very popular and flexible *family of models* is **neural networks**:



(A) Neural Networks

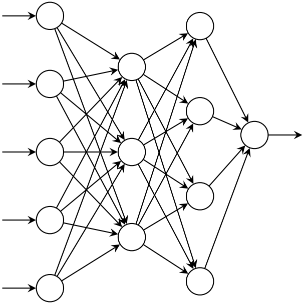
A very popular and flexible *family of models* is **neural networks**:



Given enough neurons, neural networks are **universal approximators**.

(A) Neural Networks

A very popular and flexible *family of models* is **neural networks**:



Given enough neurons, neural networks are **universal approximators**.

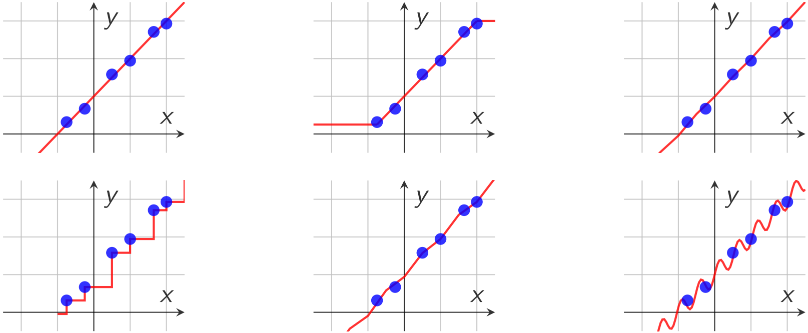
(There is large *architectural* freedom in how to organize nodes, connections and layers.)

(A) Complex Models

Any model can (in theory) be approximated by a **neural network**:

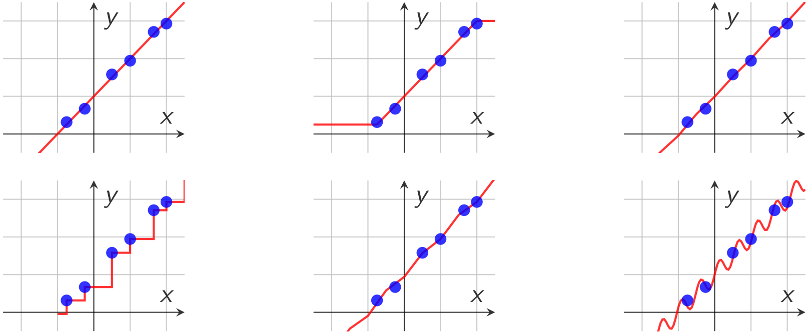
(A) Complex Models

Any model can (in theory) be approximated by a **neural network**:



(A) Complex Models

Any model can (in theory) be approximated by a **neural network**:



Scaling neural networks has been shown to be very effective for modelling *complex phenomena*.

(Aside: Theoretical vs Practical Feasibility)

The *universal approximation theorem* states that any function may be reproduced to any degree of approximation by a sufficiently large *neural network*.

(Aside: Theoretical vs Practical Feasibility)

The *universal approximation theorem* states that any function may be reproduced to any degree of approximation by a sufficiently large *neural network*.

- In **theory** we know it is feasible;

(Aside: Theoretical vs Practical Feasibility)

The *universal approximation theorem* states that any function may be reproduced to any degree of approximation by a sufficiently large *neural network*.

- In **theory** we know it is feasible;
- In **practice** there are some challenges:
  - We do not know *how large* is *sufficiently large*;
  - Different architectures may have different requirements of size.

(A) Constraints

In *neural networks*, we might still control the learning process using **prior knowledge** in the form of:

(A) Constraints

In *neural networks*, we might still control the learning process using **prior knowledge** in the form of:

- *Constraints:* forbidding some outcomes we know to be wrong;

(A) Constraints

In *neural networks*, we might still control the learning process using **prior knowledge** in the form of:

- *Constraints:* forbidding some outcomes we know to be wrong;
- *Regularizers:* penalizing some outcomes we know to be less likely;

(A) Constraints

In *neural networks*, we might still control the learning process using **prior knowledge** in the form of:

- *Constraints:* forbidding some outcomes we know to be wrong;
- *Regularizers:* penalizing some outcomes we know to be less likely;
- *Invariances:* enforcing some patterns we know to be present:

(A) Constraints

In *neural networks*, we might still control the learning process using **prior knowledge** in the form of:

- *Constraints:* forbidding some outcomes we know to be wrong;
- *Regularizers:* penalizing some outcomes we know to be less likely;
- *Invariances:* enforcing some patterns we know to be present:

(A) Constraints

In *neural networks*, we might still control the learning process using **prior knowledge** in the form of:

- *Constraints:* forbidding some outcomes we know to be wrong;
- *Regularizers:* penalizing some outcomes we know to be less likely;
- *Invariances:* enforcing some patterns we know to be present:
    - *Rotational invariances* for images;

(A) Constraints

In *neural networks*, we might still control the learning process using **prior knowledge** in the form of:

- *Constraints:* forbidding some outcomes we know to be wrong;
- *Regularizers:* penalizing some outcomes we know to be less likely;
- *Invariances:* enforcing some patterns we know to be present:
  - *Rotational invariances* for images;
  - *Preservation of objects* for videos;

(A) Constraints

In *neural networks*, we might still control the learning process using **prior knowledge** in the form of:

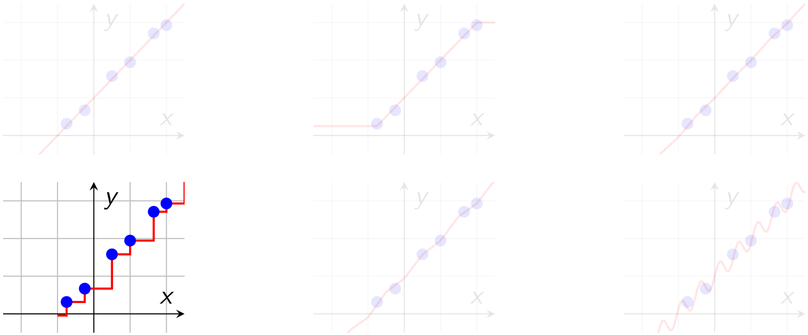
- *Constraints:* forbidding some outcomes we know to be wrong;
- *Regularizers:* penalizing some outcomes we know to be less likely;
- *Invariances:* enforcing some patterns we know to be present:
  - *Rotational invariances* for images;
  - *Preservation of objects* for videos;
  - *Context dependence* for language;

(A) Models with constraints

Constraints may help us find a meaningful model:

(A) Models with constraints

Constraints may help us find a meaningful model:



(Such a model may capture our knowledge of a discrete phenomenon).

(A) Constraints

By imposing **constraints** we:

- impose restrictions on *complexity*;
- restrict the *solution space*.

(A) Controlling Complexity

Choosing a *family of models* and imposing *constraints* allows us to balance between:

- *Underfitting:* not interpolating between observations because the model is too simple;
- *Overfitting:* adding too much artificial complexity to fit all observations.

(B) Objective and Cost Function

A different perspective on *controlling learning* is:

(B) Objective and Cost Function

A different perspective on *controlling learning* is:

- Do we know something about the shape of the models we want to learn? (**family and constraint**)

(B) Objective and Cost Function

A different perspective on *controlling learning* is:

- × Do we know something about the shape of the models we want to learn? (**family and constraint**)
- ✓ Can we measure how good a solution is? (**objective/loss**)

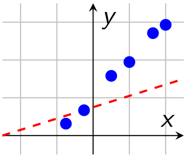
(B) Fitting to Objective

We approach the best model by *scoring* each option.

(B) Fitting to Objective

We approach the best model by *scoring* each option.

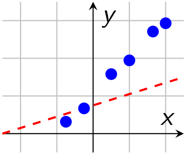
Obj: 5



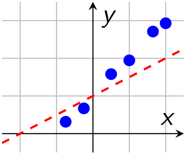
(B) Fitting to Objective

We approach the best model by *scoring* each option.

Obj: 5



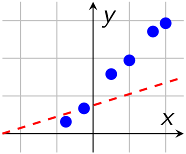
Obj: 10



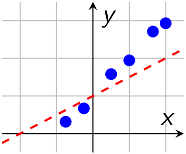
(B) Fitting to Objective

We approach the best model by *scoring* each option.

Obj: 5



Obj: 10

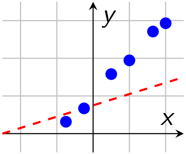


...

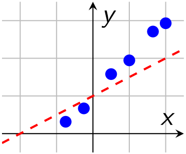
(B) Fitting to Objective

We approach the best model by *scoring* each option.

Obj: 5

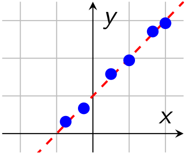


Obj: 10



...

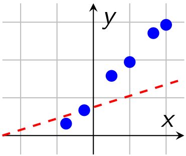
Obj: 100



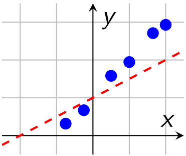
(B) Fitting to Objective

We approach the best model by *scoring* each option.

Obj: 5

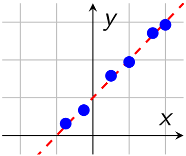


Obj: 10



...

Obj: 100



This is useful also to guide the **induction process**.

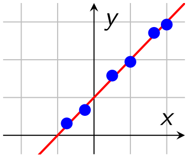
(B) Fitting to Objective

Any model could then be *ranked*:

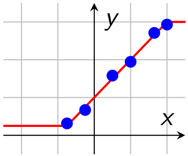
(B) Fitting to Objective

Any model could then be *ranked*:

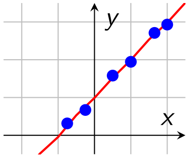
Obj: 100



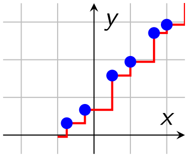
Obj: 101



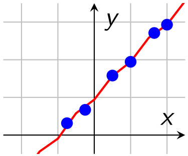
Obj: 98



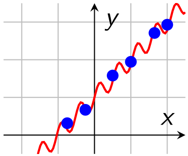
Obj: 95



Obj: 99



Obj: 92



(B) Evaluating a Model

We could use the objective to *evaluate the final induction* of a model.

(B) Evaluating a Model

We could use the objective to *evaluate the final induction* of a model.

However, we divide **datapoints** between:

- *Training data:* data used to train the model.
- *Test data:* data used to evaluate the model.

(B) Evaluating a Model

We could use the objective to *evaluate the final induction* of a model.

However, we divide **datapoints** between:

- *Training data:* data used to train the model.
- *Test data:* data used to evaluate the model.

This separation must be **rigid**.

(B) Evaluating Induction

Separation must be made and held:

(B) Evaluating Induction

Separation must be made and held:

- *Exam principle:* thou shalt not test a student on the same exercises presented in class.

(B) Evaluating Induction

Separation must be made and held:

- *Exam principle:* thou shalt not test a student on the same exercises presented in class.
- *No pre-exam cheating principle:* thou shalt not allow a student to have a peek at the exam beforehand.

Research and ML algorithms

An understanding of the internals of ML allows us a better use of ML algorithms:

Prior knowledge:

Research and ML algorithms

An understanding of the internals of ML allows us a better use of ML algorithms:

Prior knowledge:

- How can existing knowledge help *informing the induction process*?
  - What *models* can be excluded?
  - What *properties* can be enforced?
  - What *relationships* should hold?

Research and ML algorithms

An understanding of the internals of ML allows us a better use of ML algorithms:

Prior knowledge:

- How can existing knowledge help *informing the induction process*?
    - What *models* can be excluded?
    - What *properties* can be enforced?
    - What *relationships* should hold?

- What constitutes *information* and what *noise*?
    - How much do we want to penalize *complexity*?

Research and ML algorithms

An understanding of the internals of ML allows us a better use of ML algorithms:

Objectives:

Research and ML algorithms

An understanding of the internals of ML allows us a better use of ML algorithms:

Objectives:

- Is there something we want to *optimize*?

Research and ML algorithms

An understanding of the internals of ML allows us a better use of ML algorithms:

Objectives:

- Is there something we want to *optimize*?
- How do we encode this *learning aim*?
  - Is the formulation sound?
  - Is the formulation complete?
  - Do we need to rely on a proxy?

Research and ML algorithms

An understanding of the internals of ML allows us a better use of ML algorithms:

Objectives:

- Is there something we want to *optimize*?
- How do we encode this *learning aim*?
  - Is the formulation sound?
  - Is the formulation complete?
  - Do we need to rely on a proxy?
- Are we measuring the *objective* correctly?
  - Are we correctly measuring induction?

6. Conclusion

ML and Research

All this has significant impact on *thinking* and *designing research* with ML:

ML and Research

All this has significant impact on *thinking* and *designing research* with ML:

- Thinking about **data** is *necessary* but *not sufficient*.

ML and Research

All this has significant impact on *thinking* and *designing research* with ML:

- Thinking about **data** is *necessary* but *not sufficient*.

One might want to think about:

ML and Research

All this has significant impact on *thinking* and *designing research* with ML:

- Thinking about **data** is *necessary* but *not sufficient*.

One might want to think about:

- What **questions** are to be asked?

ML and Research

All this has significant impact on *thinking* and *designing research* with ML:

- Thinking about **data** is *necessary* but *not sufficient*.

One might want to think about:

- What **questions** are to be asked?
- What **assumptions** can we make?

ML and Research

All this has significant impact on *thinking* and *designing research* with ML:

- Thinking about **data** is *necessary* but *not sufficient*.

One might want to think about:

- What **questions** are to be asked?
- What **assumptions** can we make?
- How **inductive procedures** can help with answering?

ML and Research

All this has significant impact on *thinking* and *designing research* with ML:

- Thinking about **data** is *necessary* but *not sufficient*.

One might want to think about:

- What **questions** are to be asked?
- What **assumptions** can we make?
- How **inductive procedures** can help with answering?
- What **prior knowledge** can we bring to the model?

ML and Research

All this has significant impact on *thinking* and *designing research* with ML:

- Thinking about **data** is *necessary* but *not sufficient*.

One might want to think about:

- What **questions** are to be asked?
- What **assumptions** can we make?
- How **inductive procedures** can help with answering?
- What **prior knowledge** can we bring to the model?
- How can we restrict the solution space with **constraints**?

ML and Research

All this has significant impact on *thinking* and *designing research* with ML:

- Thinking about **data** is *necessary* but *not sufficient*.

One might want to think about:

- What **questions** are to be asked?
- What **assumptions** can we make?
- How **inductive procedures** can help with answering?
- What **prior knowledge** can we bring to the model?
- How can we restrict the solution space with **constraints**?
- What **objective** better describes our aims?

ML and Research

Although much can be learned just from *raw data*, it is very doubtful we want to do it.

ML and Research

Although much can be learned just from *raw data*, it is very doubtful we want to do it.

Properly designing learning and providing prior knowledge:

ML and Research

Although much can be learned just from *raw data*, it is very doubtful we want to do it.

Properly designing learning and providing prior knowledge:

- ✓ Less computational resources

ML and Research

Although much can be learned just from *raw data*, it is very doubtful we want to do it.

Properly designing learning and providing prior knowledge:

- ✓ Less computational resources
- ✓ Less time

ML and Research

Although much can be learned just from *raw data*, it is very doubtful we want to do it.

Properly designing learning and providing prior knowledge:

- ✓ Less computational resources
- ✓ Less time
- ✓ Less energy

ML and Research

Although much can be learned just from *raw data*, it is very doubtful we want to do it.

Properly designing learning and providing prior knowledge:

- ✓ Less computational resources
- ✓ Less time
- ✓ Less energy
- ✓ More interpretable

ML and Research

Although much can be learned just from *raw data*, it is very doubtful we want to do it.

Properly designing learning and providing prior knowledge:

- ✓ Less computational resources
- ✓ Less time
- ✓ Less energy
- ✓ More interpretable

- ? Fewer surprises (in positive and negative sense).

ML and Research

Although much can be learned just from *raw data*, it is very doubtful we want to do it.

Properly designing learning and providing prior knowledge:

- ✓ Less computational resources
- ✓ Less time
- ✓ Less energy
- ✓ More interpretable

? Fewer surprises (in positive and negative sense).

AlphaGo vs AlphaZero

Some Extensions to Standard Machine Learning

Fabio Massimo Zennaro

University of Bergen

7. ML Extensions

Extensions

Standard ML is grounded on a set of assumptions and requirements for working properly.

Extensions

Standard ML is grounded on a set of assumptions and requirements for working properly.

Different *subfields* explore realistic extensions:

Extensions

Standard ML is grounded on a set of assumptions and requirements for working properly.

Different *subfields* explore realistic extensions:

- Heterogeneous, incomplete, missing data:
  - *Feature engineering*

Extensions

Standard ML is grounded on a set of assumptions and requirements for working properly.

Different *subfields* explore realistic extensions:

- Heterogeneous, incomplete, missing data:
  - *Feature engineering*
- Data coming from different generating processes/distributions:
  - *Transfer learning*, *Covariate Shift Adaptation*

Extensions

Standard ML is grounded on a set of assumptions and requirements for working properly.

Different *subfields* explore realistic extensions:

- Heterogeneous, incomplete, missing data:
  - *Feature engineering*
- Data coming from different generating processes/distributions:
  - *Transfer learning*, *Covariate Shift Adaptation*
- Data coming from a generating process changing in time:
  - *Adaptive learning*, *Time-variant models*

Extensions

Standard ML is grounded on a set of assumptions and requirements for working properly.

Different *subfields* explore realistic extensions:

- Heterogeneous, incomplete, missing data:
  - *Feature engineering*
- Data coming from different generating processes/distributions:
  - *Transfer learning*, *Covariate Shift Adaptation*
- Data coming from a generating process changing in time:
  - *Adaptive learning*, *Time-variant models*
- Unbalanced data:
  - *Weighted learning, Cost-sensitive learning*

Extensions

Standard ML is grounded on a set of assumptions and requirements for working properly.

Different *subfields* explore realistic extensions:

- Heterogeneous, incomplete, missing data:
  - *Feature engineering*
- Data coming from different generating processes/distributions:
  - *Transfer learning*, *Covariate Shift Adaptation*
- Data coming from a generating process changing in time:
  - *Adaptive learning*, *Time-variant models*
- Unbalanced data:
  - *Weighted learning, Cost-sensitive learning*
- Focus on decision-making:
  - *Reinforcement learning, Cost-sensitive learning*

Extensions

Different *subfields* explore realistic extensions:

Extensions

Different *subfields* explore realistic extensions:

- Deciding which data to acquire:
    - *Active learning*

Extensions

Different *subfields* explore realistic extensions:

- Deciding which data to acquire:
  - *Active learning*
- Availability of multiple models:
  - *Ensemble learning*

Extensions

Different *subfields* explore realistic extensions:

- Deciding which data to acquire:
  - *Active learning*
- Availability of multiple models:
  - *Ensemble learning*
- Interest in probabilities:
  - *Bayesian learning*, *Uncertainty quantification*

Extensions

Different *subfields* explore realistic extensions:

- Deciding which data to acquire:
  - *Active learning*
- Availability of multiple models:
  - *Ensemble learning*
- Interest in probabilities:
  - *Bayesian learning*, *Uncertainty quantification*
- Interest in explainable methods:
  - *XAI*

Extensions

Different *subfields* explore realistic extensions:

- Deciding which data to acquire:
  - *Active learning*
- Availability of multiple models:
  - *Ensemble learning*
- Interest in probabilities:
  - *Bayesian learning*, *Uncertainty quantification*
- Interest in explainable methods:
  - *XAI*
- Integration with other reasoning modalities:
  - *Neuro-symbolic learning*

Extensions

Different *subfields* explore realistic extensions:

- Deciding which data to acquire:
    - *Active learning*
- Availability of multiple models:
    - *Ensemble learning*
- Interest in probabilities:
    - *Bayesian learning*, *Uncertainty quantification*
- Interest in explainable methods:
    - *XAI*
- Integration with other reasoning modalities:
    - *Neuro-symbolic learning*
- Interest in causality:
    - *Causal learning*

Causal Machine Learning

Fabio Massimo Zennaro

University of Bergen

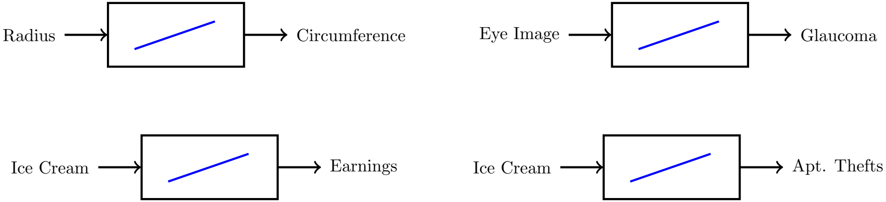
8. Causality

Prediction

We have seen a number of models that are good at **prediction**.

Prediction

We have seen a number of models that are good at **prediction**.



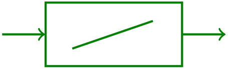
Control

What about **control**?

Control

What about **control**?

Radius



Circumference

Control

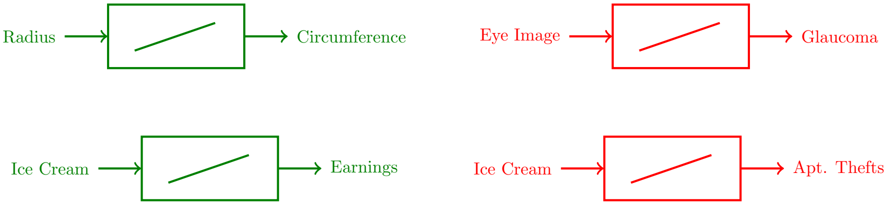
What about **control**?

Control

What about **control**?

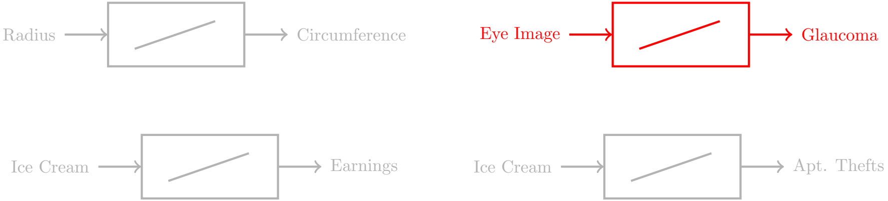
Control

What about **control**?



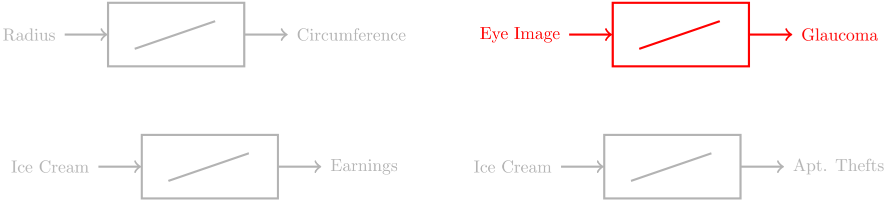
Failure of Control

We may *fail to control* for different reasons.



Failure of Control

We may *fail to control* for different reasons.



Sometime our models may be **anti-causal**.

Failure of Control

We may *fail to control* for different reasons.

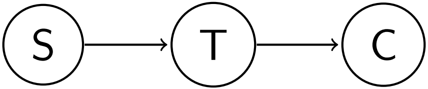
Failure of Control

We may *fail to control* for different reasons.

Sometime our models may ignore **common causes**/**confounders**.

Structural Causal Models

We use a **graphical language** to express causal relations:



Structural Causal Models

We use a **graphical language** to express causal relations:

$$S \longrightarrow T \longrightarrow C$$

We use **structural causal models** (SCM):

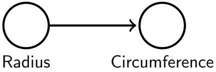
- to *compute relations* between variables;
- to *reason causally* beyond pure statistic-correlation.

Examples

We can model positive examples with SCMs:

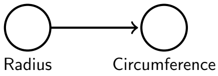
Examples

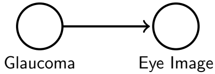
We can model positive examples with SCMs:



Examples

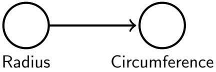
We can model positive examples with SCMs:

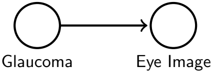


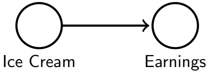


Examples

We can model positive examples with SCMs:

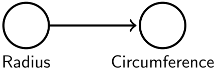


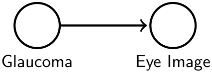


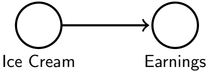


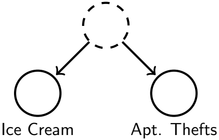
Examples

We can model positive examples with SCMs:



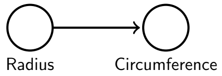


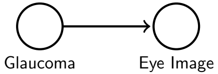


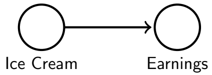


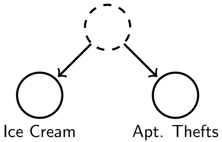
Examples

We can model positive examples with SCMs:









All these models are example of **structural prior knowledge**.

Interventions

SCMs formalize the idea of control through **interventions**:

$$S \longrightarrow T \longrightarrow C$$

Interventions

SCMs formalize the idea of control through **interventions**:

$$S \longrightarrow T \longrightarrow C$$

Intervention $do(T = 1)$:

Interventions

SCMs formalize the idea of control through **interventions**:

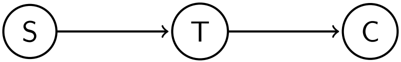
$$S \rightarrow T \rightarrow C$$

Intervention $do(T = 1)$:

1. Removes incoming edges in the intervened node

Interventions

SCMs formalize the idea of control through **interventions**:

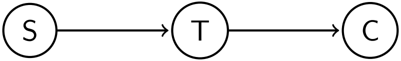


Intervention $do(T = 1)$:

1. Removes incoming edges in the intervened node
2. Sets the value of the intervened node

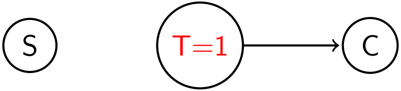
Interventions

SCMs formalize the idea of control through **interventions**:



Intervention $do(T = 1)$:

1. Removes incoming edges in the intervened node
2. Sets the value of the intervened node



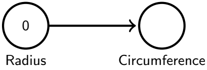
Examples

We can model interventions on positive examples:

Examples

We can model interventions on positive examples:

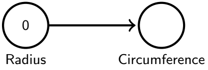
do(Radius=0)



Examples

We can model interventions on positive examples:

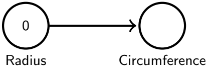
do(Radius=0)



Examples

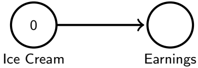
We can model interventions on positive examples:

do(Radius=0)



do(Eye Image=0)

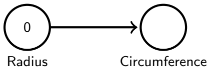
do(Ice Cream=0)



Examples

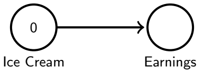
We can model interventions on positive examples:

do(Radius=0)

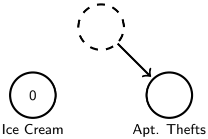


do(Eye Image=0)

do(Ice Cream=0)



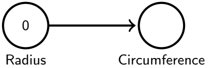
do(Ice Cream=0)



Examples

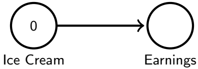
We can model interventions on positive examples:

do(Radius=0)

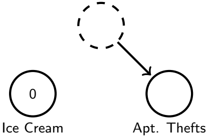


do(Eye Image=0)

do(Ice Cream=0)



do(Ice Cream=0)



Interventions change the **structure** of a model and the relation between variables.

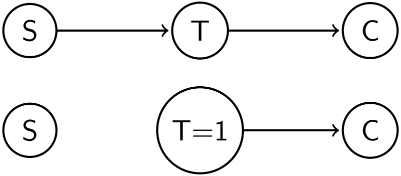
Causal Inference

Causal inference allows us to relate *intervened models*:

$$S \longrightarrow T \longrightarrow C$$

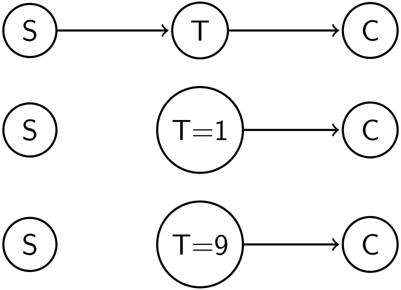
Causal Inference

Causal inference allows us to relate *intervened models*:



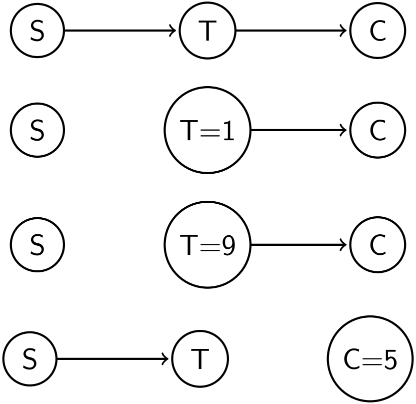
Causal Inference

Causal inference allows us to relate *intervened models*:



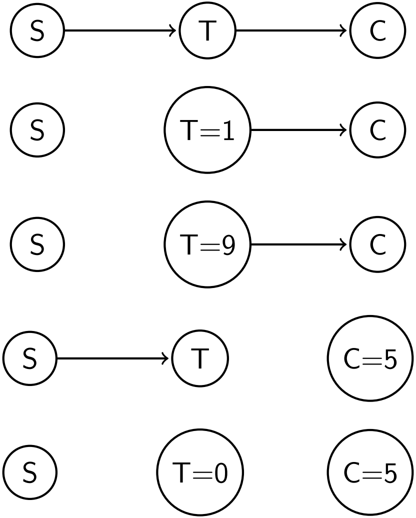
Causal Inference

Causal inference allows us to relate *intervened models*:



Causal Inference

Causal inference allows us to relate *intervened models*:



Research and causality

"All the impressive achievements of deep learning amount to just curve fitting." (Judea Pearl)

Research and causality

"All the impressive achievements of deep learning amount to just curve fitting." (Judea Pearl)

Causality has *something more* to offer than standard ML.

Research and causality

> *"All the impressive achievements of deep learning amount to just curve fitting." (Judea Pearl)*

Causality has *something more* to offer than standard ML.

We *extend* the **domain** of the models through causal reasoning.

- From each *base*/*observational* model we can generate multiple *interventional* models!

Research and causality

Causal modelling would be relevant:

Research and causality

Causal modelling would be relevant:

- If we want to *explain* and *describe* a system;

Research and causality

Causal modelling would be relevant:

- If we want to *explain* and *describe* a system;
- If we want to *control* a system;

Research and causality

Causal modelling would be relevant:

- If we want to *explain* and *describe* a system;
- If we want to *control* a system;
- If we want to *exploit* data from different regimes;

Research and causality

Causal modelling would be relevant:

- If we want to *explain* and *describe* a system;
- If we want to *control* a system;
- If we want to *exploit* data from different regimes;
- If we want to model *different environments* through interventions.

Thanks!

Thank you for listening!