Machine Learning Tutorial
I - Introduction

ADA IX Summer School
20-22 May 2018, Valencia - Spain

Emille E. O. Ishida
Laboratoire de Physique de Clermont - Université Clermont-Auvergne
Clermont Ferrand, France

In collaboration with
Alexandre Boucaud
Paris-Saclay Center for Data Science

Emille E. O. Ishida

Researcher at the Laboratoire de Physique de Clermont

Clermont Ferrand, France

Type Ia Supernova + Machine Learning + Bayesian Statistics

Interdisciplinary science development

Co-chair of the Cosmostatistics Initiative (COIN)

CNRS - MOMENTUM Laureate

Clermont Ferrand, France

COIN
Cosmostatistics Initiative

UNIVERSITÉ Clermont Auvergne

Truffade

Large Synoptic Survey Telescope

ZTF

Puy-de-Dome

Summary

I. Introduction

II. Practical examples

III. From NN to CNN by *Alexandre Boucaud*

IV. Data Science Workflow by *Alexandre Boucaud*

V: Beyond textbook Machine Learning

Acknowledgments

Adam Miller

Andrew Ng

Fabian Gieseke

David Hogg

David Kirby

Daniela Huppenkothen

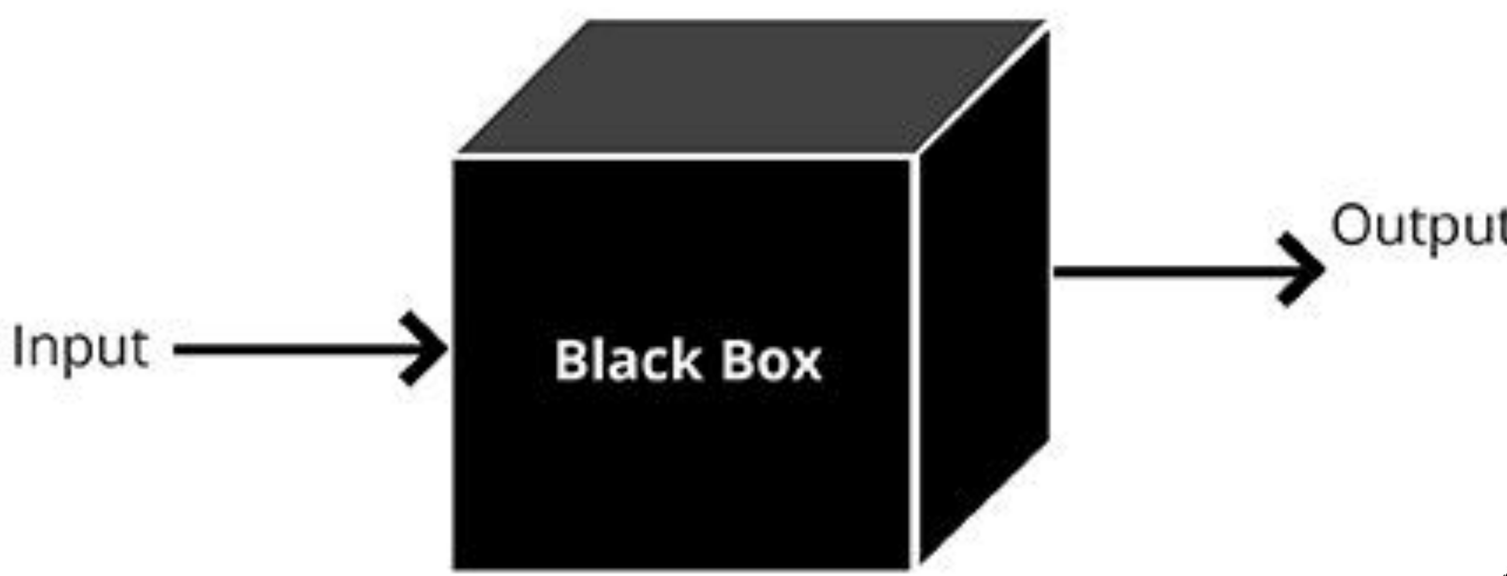
Jake VanderPlas

Ricardo Vilalta

The Cosmostatistics Initiative

Disclaimer 1

Beware of Black Boxes!!!



Disclaimer 2

for now ...

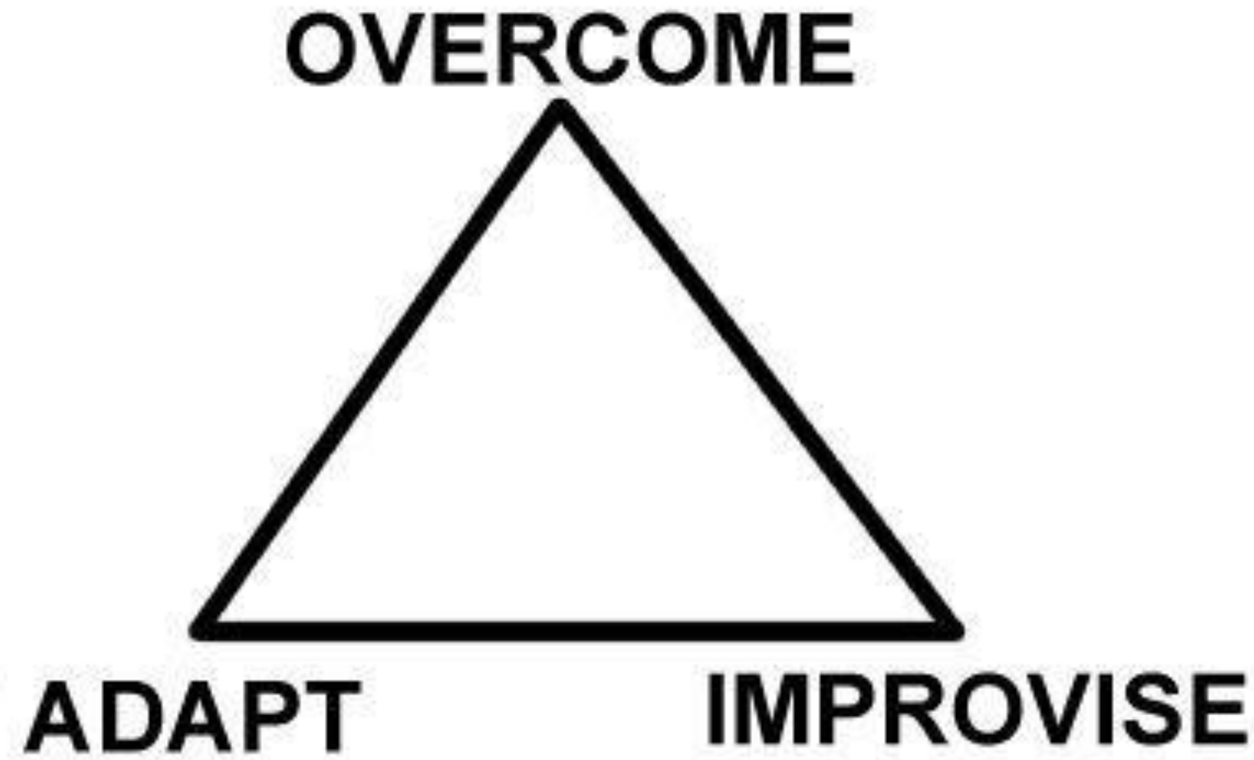
What is learning ?

What is learning ?

“A relatively permanent change in behaviour due to past experiences.”

D. Coon, *Introduction to psychology: exploration and application* (1983)

Learning *or the Power to Adapt*



What is Machine Learning ?

What is Machine Learning ?

“ML is a field of computer science that gives computer systems the ability to learn with data, without being explicitly programmed.”

Samuel, Arthur (1959), *Some Studies in Machine Learning Using the Game of Checkers,* IBM JRD 3 (3)

What is Machine Learning ?

"ML is a field of computer science that gives computer systems the ability to learn with data, ***without being explicitly programmed****."*

Samuel, Arthur (1959), *Some Studies in Machine Learning Using the Game of Checkers,* IBM JRD 3 (3)

What is Machine Learning ?

WIKIPEDIA
The Free Encyclopedia

*ML is a field of computer science that gives computer systems the ability to "learn" **(i.e., progressively improve performance on a specific task)** with data, without being explicitly programmed.*

Samuel, Arthur (1959), *Some Studies in Machine Learning Using the Game of Checkers,* IBM JRD 3 (3)

Machine Learning: a definition

"A computer program L 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."

Mitchell, T. (1997). *Machine Learning*. McGraw Hill. p. 2

Machine Learning: a definition

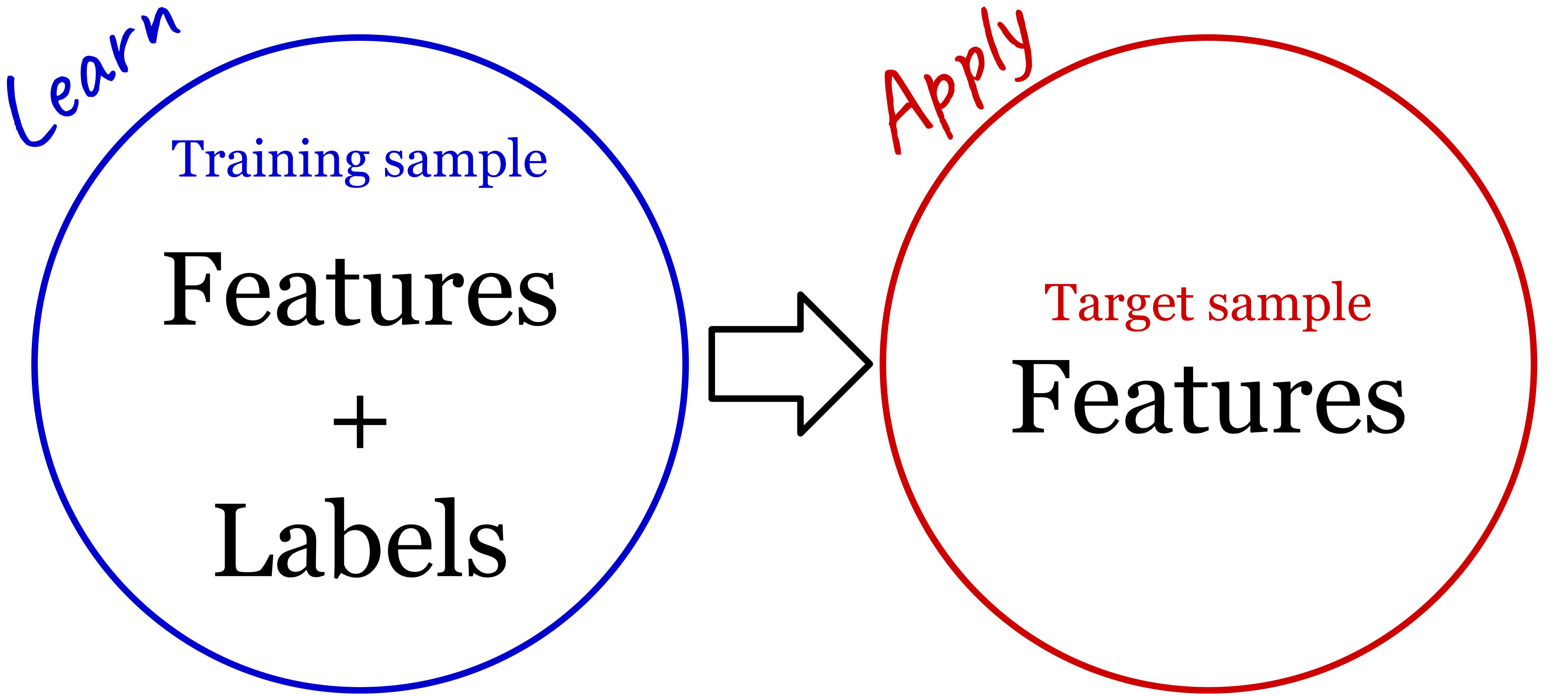
*"A computer program L is said to learn from **experience E** with respect to some class of **tasks T** and **performance measure P** if its performance at tasks in T, as measured by P, improves with experience E."*

Mitchell, T. (1997). *Machine Learning*. McGraw Hill. p. 2

Categories of Machine Learning:

Supervised Learning

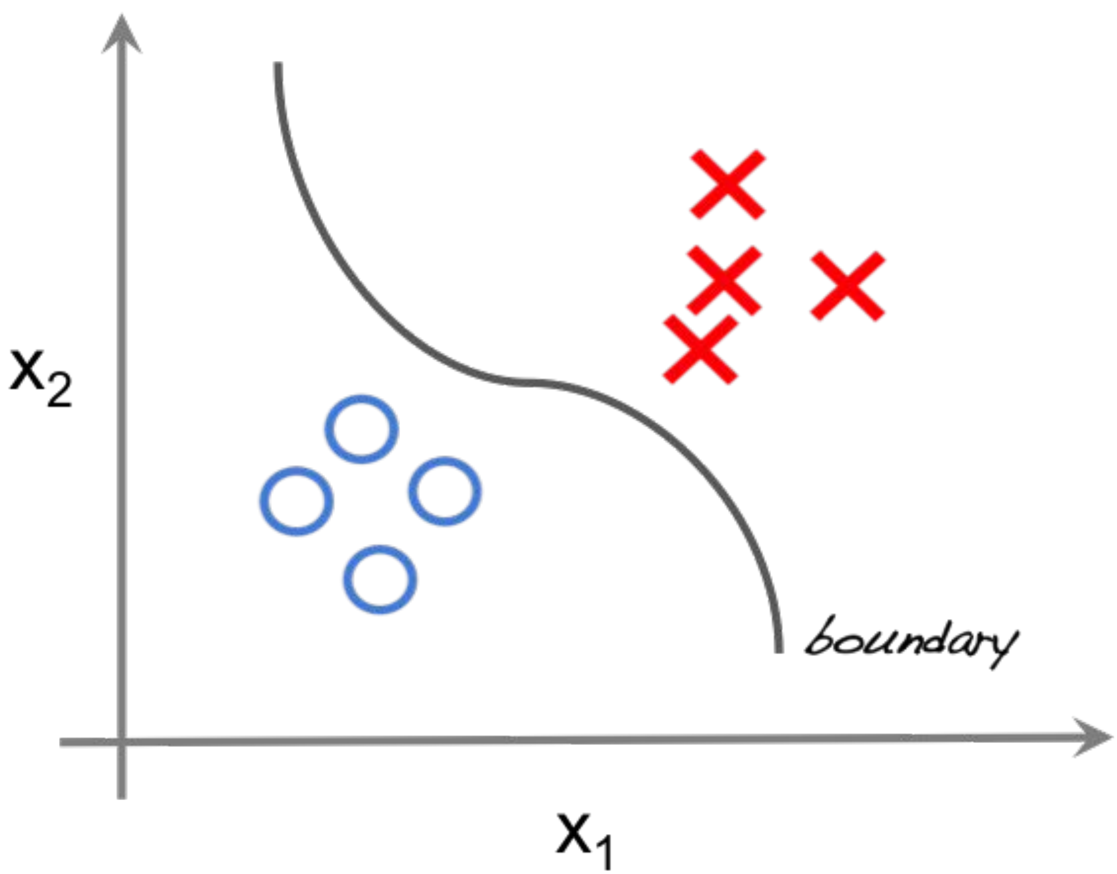
Learn by example



Categories of Machine Learning:

Supervised Learning

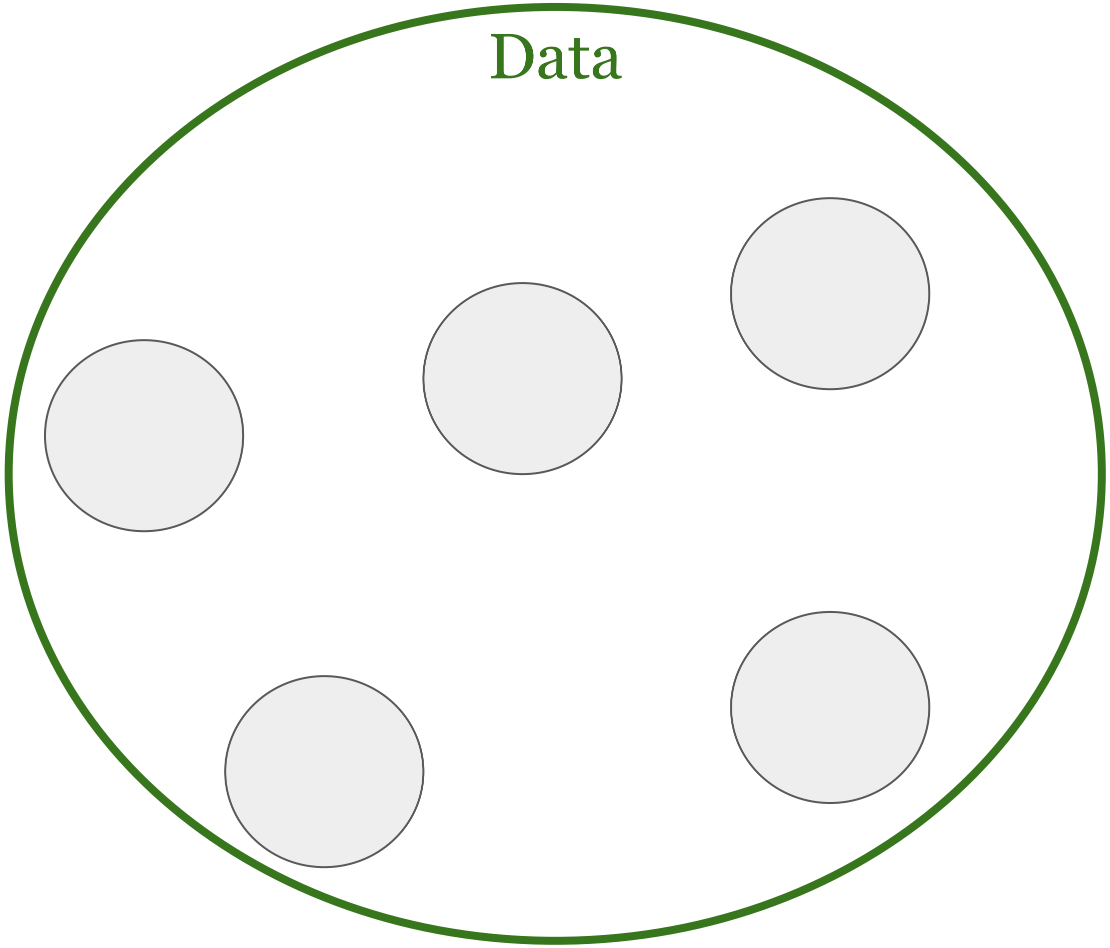
Learn by example



Categories of Machine Learning:

Unsupervised Learning

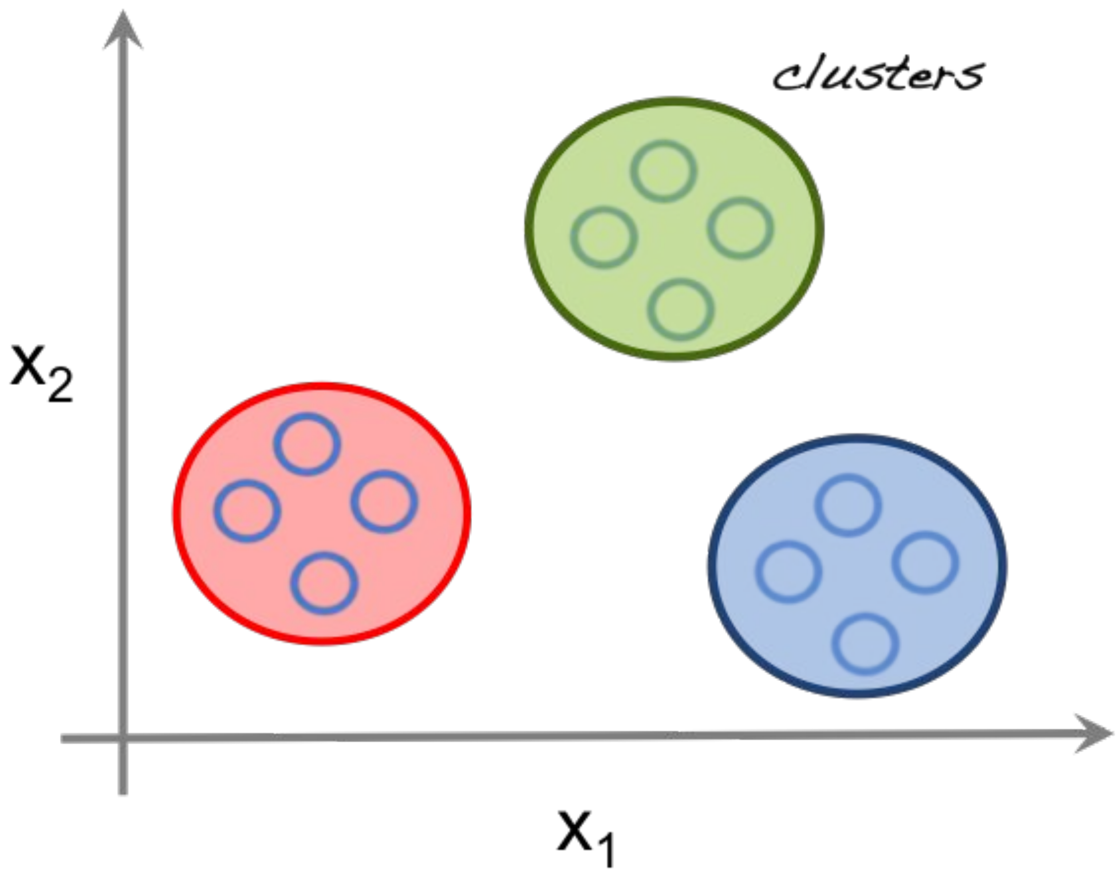
Search for data structures



Categories of Machine Learning:

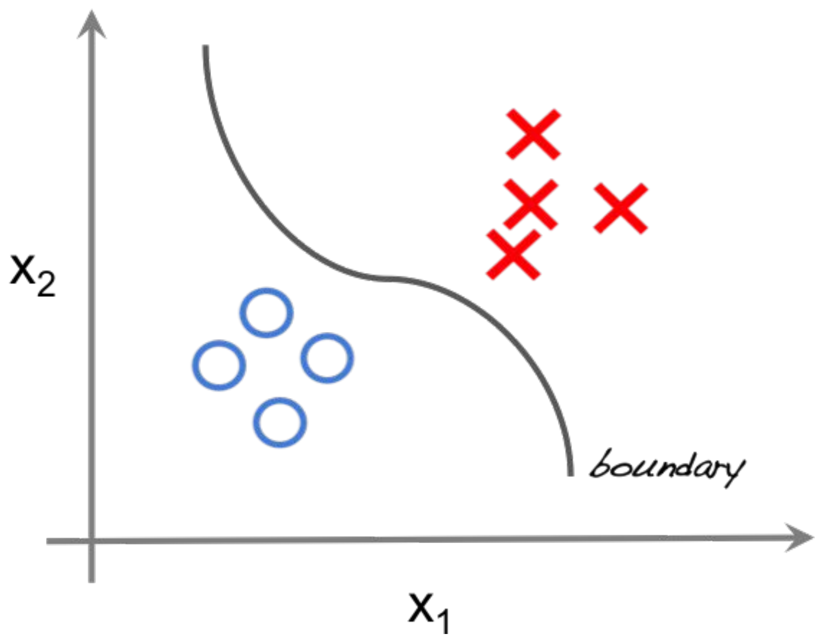
Unsupervised Learning

Search for data structures



Categories of Machine Learning:

Supervised x Unsupervised

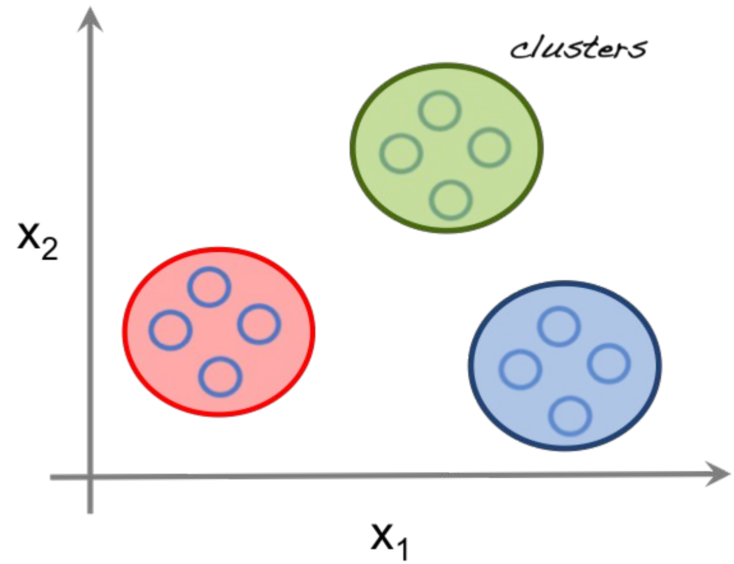


Tranining sample:

features + labels

Target sample:

features



Data sample:

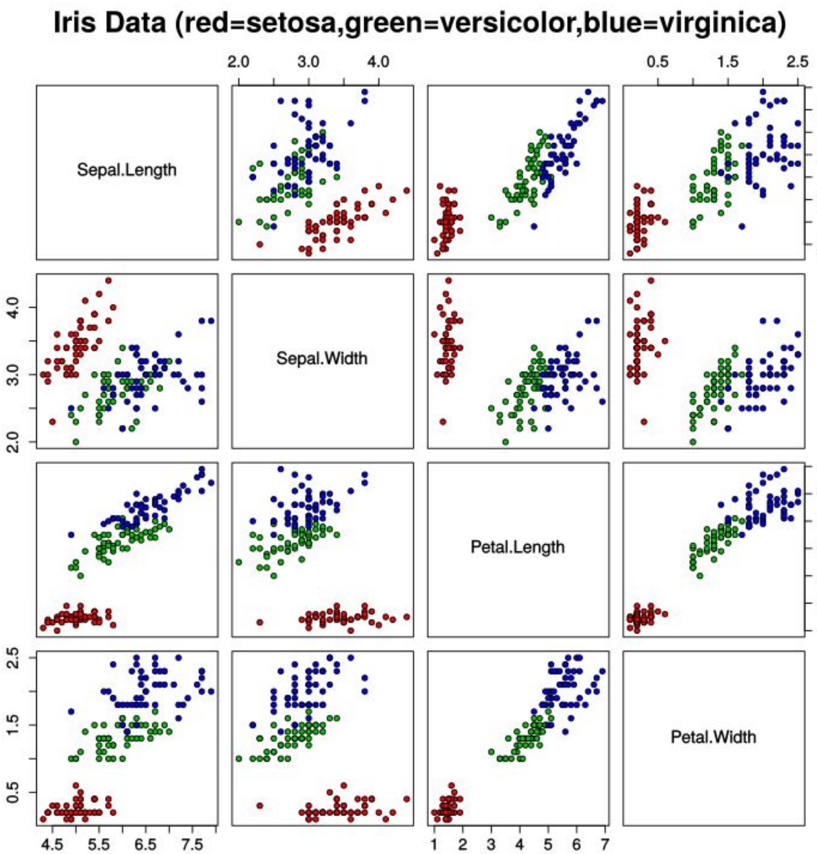
features

Supervised Learning

Example 1: the Iris dataset

Presented by R. A. Fisher (1936)

50 samples of each class



Setosa

Versicolor

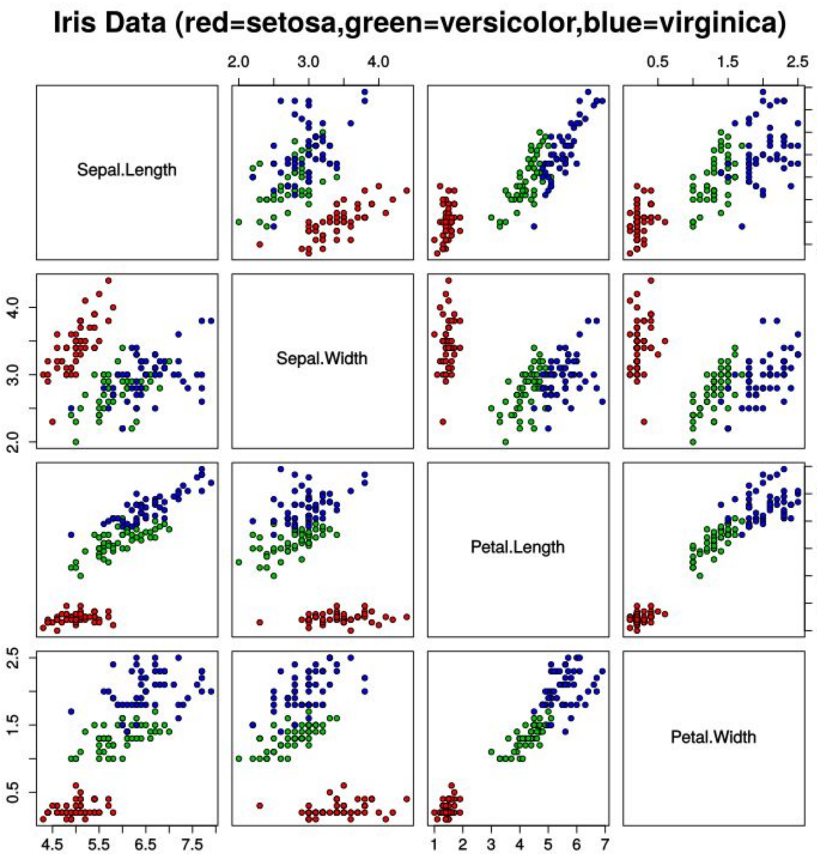
Virginica

https://en.wikipedia.org/wiki/Iris_flower_data_set

Example 1: the Iris dataset

Presented by R. A. Fisher (1936)

50 samples of each class



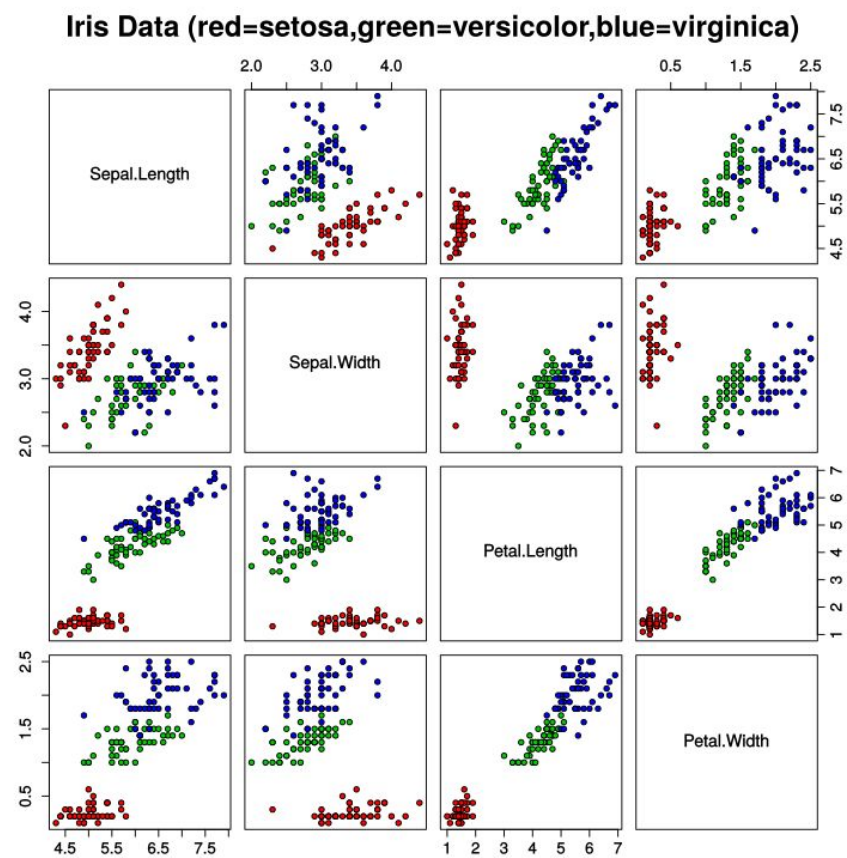
experience E	→	data
task T	→	problem
performance P	→	metric
program L	→	algorithm

https://en.wikipedia.org/wiki/Iris_flower_data_set

Example 1: the Iris dataset

Presented by R. A. Fisher (1936)

50 samples of each class



https://en.wikipedia.org/wiki/Iris_flower_data_set

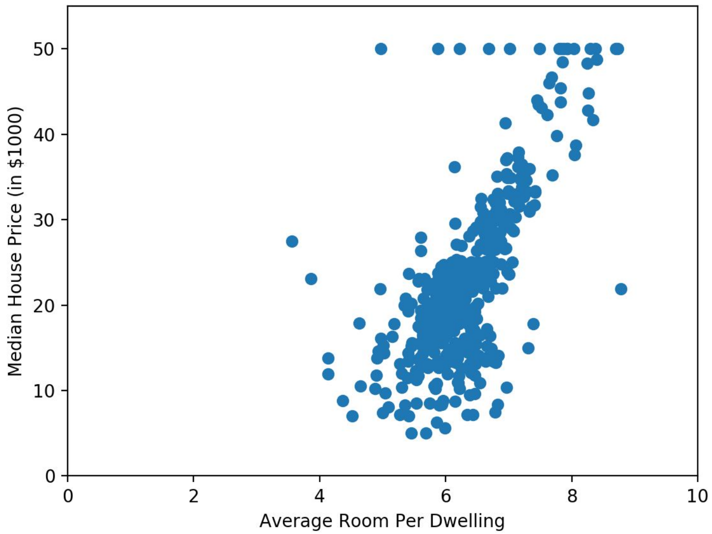
experience E	→	data
task T	→	class.
performance P	→	accuracy
program L	→	boundary

Does L learn?

Example 2: the Boston dataset

Housing prices in Boston area

506 samples and 14 features



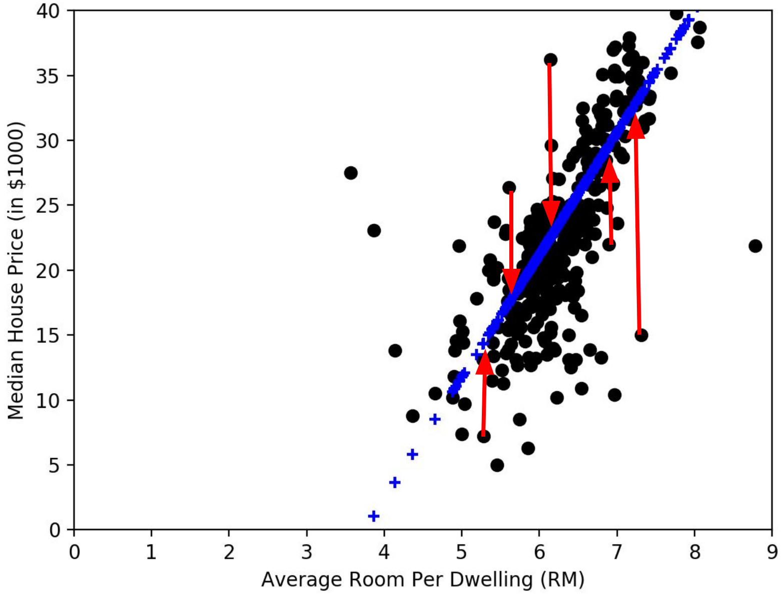
experience E	→	data
task T	→	problem
performance P	→	metric
program L	→	algorithm

Harrison, D. and Rubinfeld, D.L. *'Hedonic prices and the demand for clean air'*, JEE & M, vol.5, 81-102, 1978

Example 2: the Boston dataset

Housing prices in Boston area

506 samples and 14 features



experience E	→	data
task T	→	reg.
performance P	→	RMS
program L	→	relation

y = ax + b

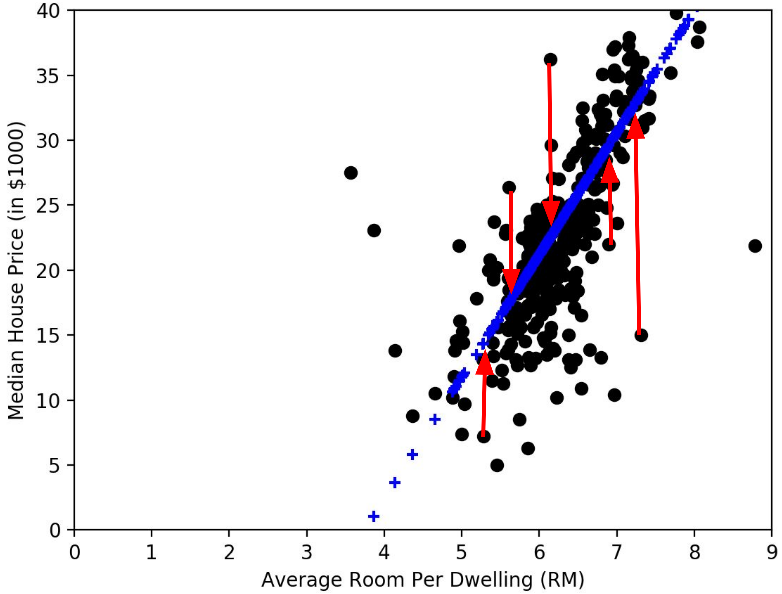
$$\{a, b\} \leftarrow \min \left[\sum_{i=1}^{N} (y_i - (ax_i + b))^2 \right]$$

Harrison, D. and Rubinfeld, D.L. *'Hedonic prices and the demand for clean air'*, JEE & M, vol.5, 81-102, 1978

Example 2: the Boston dataset

Housing prices in Boston area

506 samples and 14 features



experience E	→	data
task T	→	reg.
performance P	→	res.^2
program L	→	relation

$$y = ax + b$$

$$\{a, b\} \leftarrow \min \left[\sum_{i=1}^{N} (y_i - (ax_i + b))^2 \right]$$

Does L learn?

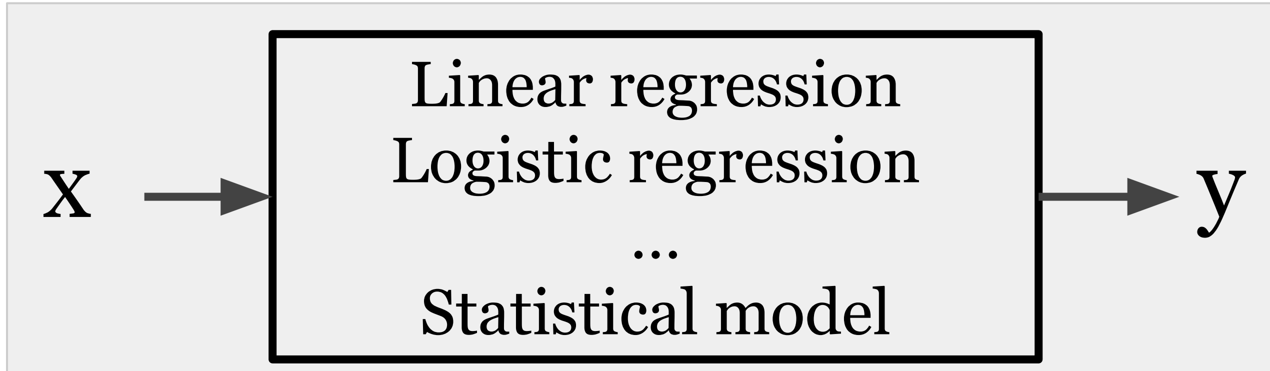
Harrison, D. and Rubinfeld, D.L. *'Hedonic prices and the demand for clean air'*, JEE & M, vol.5, 81-102, 1978

Hypothesis:

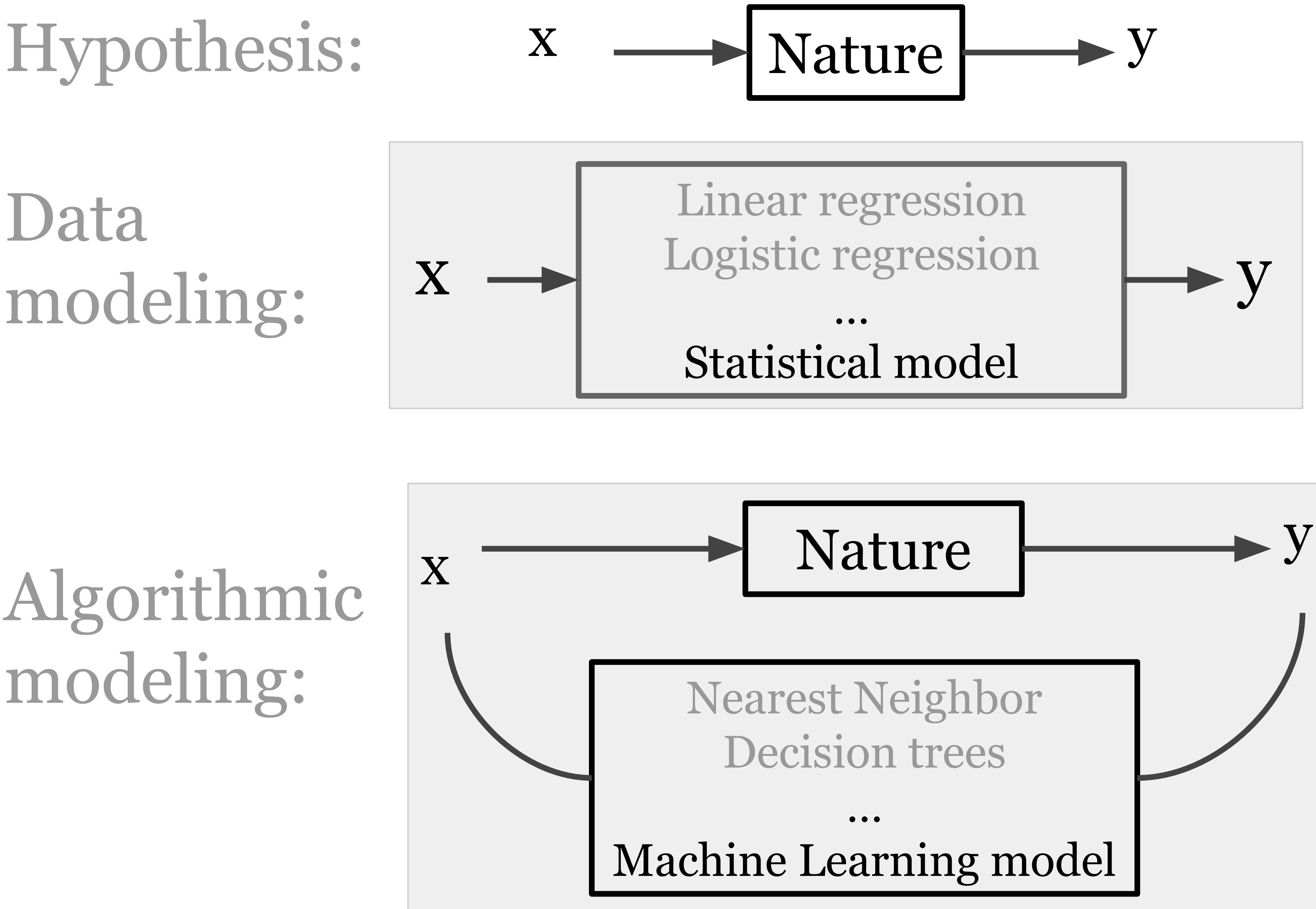
Breiman, L., *Statistical Modeling: The Two Cultures*, Stat. SciVolume 16 (2001)

Hypothesis:

Data modeling:



Breiman, L., *Statistical Modeling: The Two Cultures*, Stat. SciVolume 16 (2001)



Breiman, L., *Statistical Modeling: The Two Cultures*, Stat. Sci, Volume 16 (2001)

Supervised ML model

data	training, target
χ	set of all samples, x
Y	set of possible labels, y
h_{train}	learner: $y_{est;i} = h_{train}(x_i)$
L	loss function

Goal: *minimize L*

Shai and Shai, *Understanding ML: From Theory to Algorithms*, 2014, CUP

Supervised ML model

data	training, target
χ	set of all samples, x
Y	set of possible labels, y
h_{train}	learner: $y_{est;i} = h_{train}(x_i)$
L	loss function

Why should this work?

Goal: *minimize L*

Shai and Shai, *Understanding ML: From Theory to Algorithms*, 2014, CUP

Supervised ML model

data	training, target
χ	set of all samples, x
Y	set of possible labels, y
h_{train}	learner: $y_{est;i} = h_{train}(x_i)$
L	Loss function

Hypothesis

Data generation model:

$\mathrm{x}_\mathrm{i} \sim P_\chi$

f true labeling function, $y_i = f(x_i)$

$L_{data,f}(h) \equiv P_{x\sim data}(h_{train}(x) \neq f(x))$

Shai and Shai, *Understanding ML: From Theory to Algorithms*, 2014, CUP

Supervised ML model

data **training**, target

χ	set of all samples, x
Y	set of possible labels, y
h_{train}	learner: $y_{est;i} = h_{train}(x_i)$
L	Loss function

Hypothesis: Data dependent

Data generation model:

$x_i \sim P_{\chi}$

f true labeling function, $y_i = f(x_i)$

$$L_{data,f}(h) \equiv P_{x \sim data}(h_{train}(x) \neq f(x))$$

Shai and Shai, *Understanding ML: From Theory to Algorithms*, 2014, CUP

Supervised ML model

We will never have access to P and f

data	training, target
χ	set of all samples, x
Y	set of possible labels, y
h_{train}	learner: $y_{est;i} = h_{train}(x_i)$
L	Loss function

Data generation model:

$x_i \sim \cancel{P_\chi}$

$\cancel{f}$ true labeling function, $y_i = f(\cancel{x_i})$

$$L_{data,f}(h) \equiv \cancel{P}_{x\sim data}(h_{train}(x) \neq \cancel{f}(x))$$

Shai and Shai, *Understanding ML: From Theory to Algorithms*, 2014, CUP

Supervised ML model

We will never have access to P and f

data training, target

χ set of all samples, x

Y set of possible labels, y

h_{train} learner: $y_{est;i} = h_{train}(x_i)$

L loss function $\leftarrow$ good approximation

Goal: *minimize L*

Shai and Shai, *Understanding ML: From Theory to Algorithms*, 2014, CUP

Supervised ML **model**

data	**training**, target
χ	set of all samples, x
Y	set of possible labels, y
h_{train}	learner: $y_{est;i} = h_{train}(x_i)$
L	Loss function

Hypothesis: Data dependent

Data generation model:

$x_i \sim P_{\chi}$

f true labeling function, $y_i = f(x_i)$

$L_{data,f}(h) \equiv P_{x\sim data}(h_{train}(x) \neq f(x))$

Shai and Shai, *Understanding ML: From Theory to Algorithms*, 2014, CUP

Supervised ML **model**

data	**training**, target
X	set of all examples, x
Y	set of possible labels, y
h_{train}	learner: $y_{est;i} = h_{train}(x_i)$
L	Loss function

Machine Learning algorithm

Hypothesis: Data dependent

Data generation model:

$x_i \sim P_X$

f true labeling function, $y_i = f(x_i)$

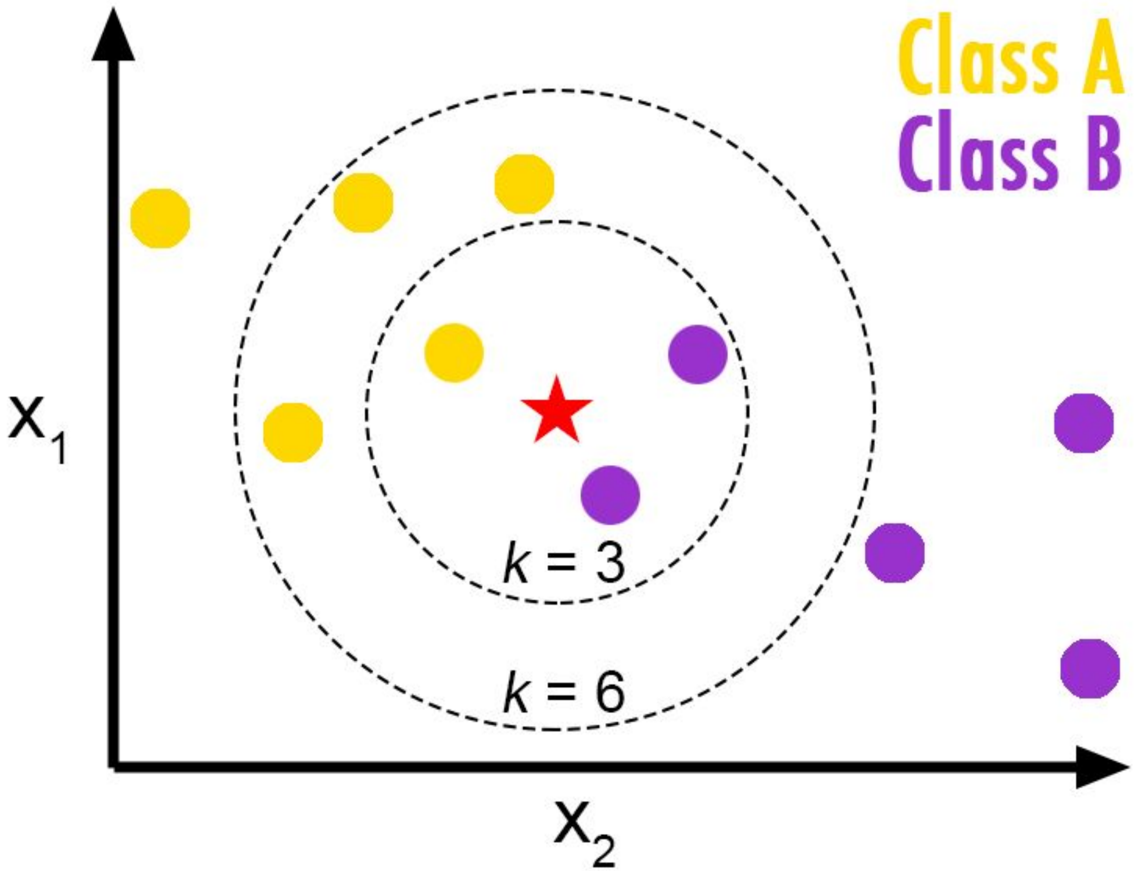
$L_{data,f}(h) \equiv P_{x \sim data}(h_{train}(x) \neq f(x))$

Shai and Shai, *Understanding ML: From Theory to Algorithms*, 2014, CUP

Example of supervised ML algorithm for classification

k-Nearest Neighbor (kNN)

Distance based

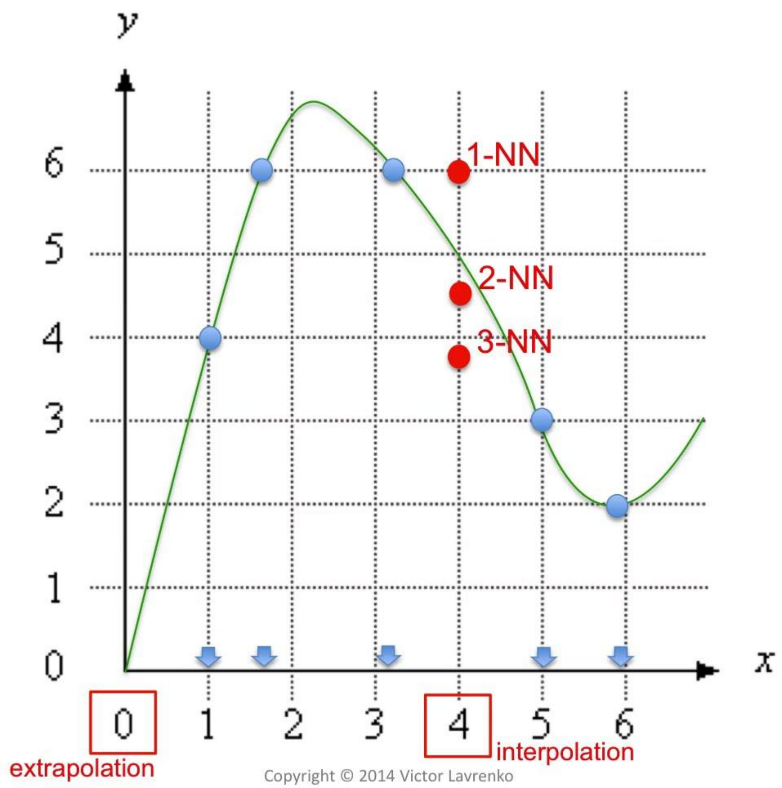


Example of supervised ML algorithm for regression

k-Nearest Neighbor (kNN)

Distance based

Example: kNN regression in 1-d



https://www.youtube.com/watch?v=3lp5CmSwrHI

Example of supervised ML algorithm for classification

Support Vector Machines (SVM)

Search for hyperplanes

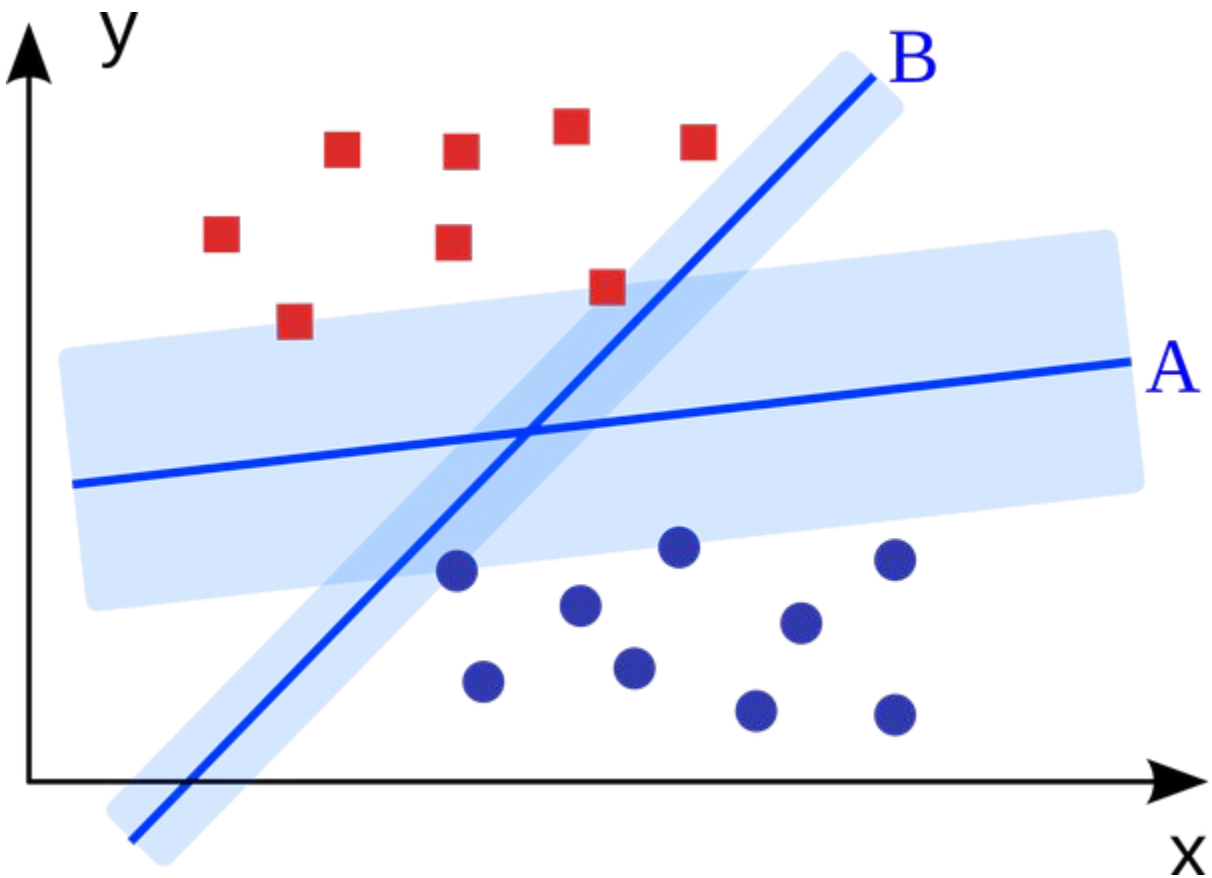
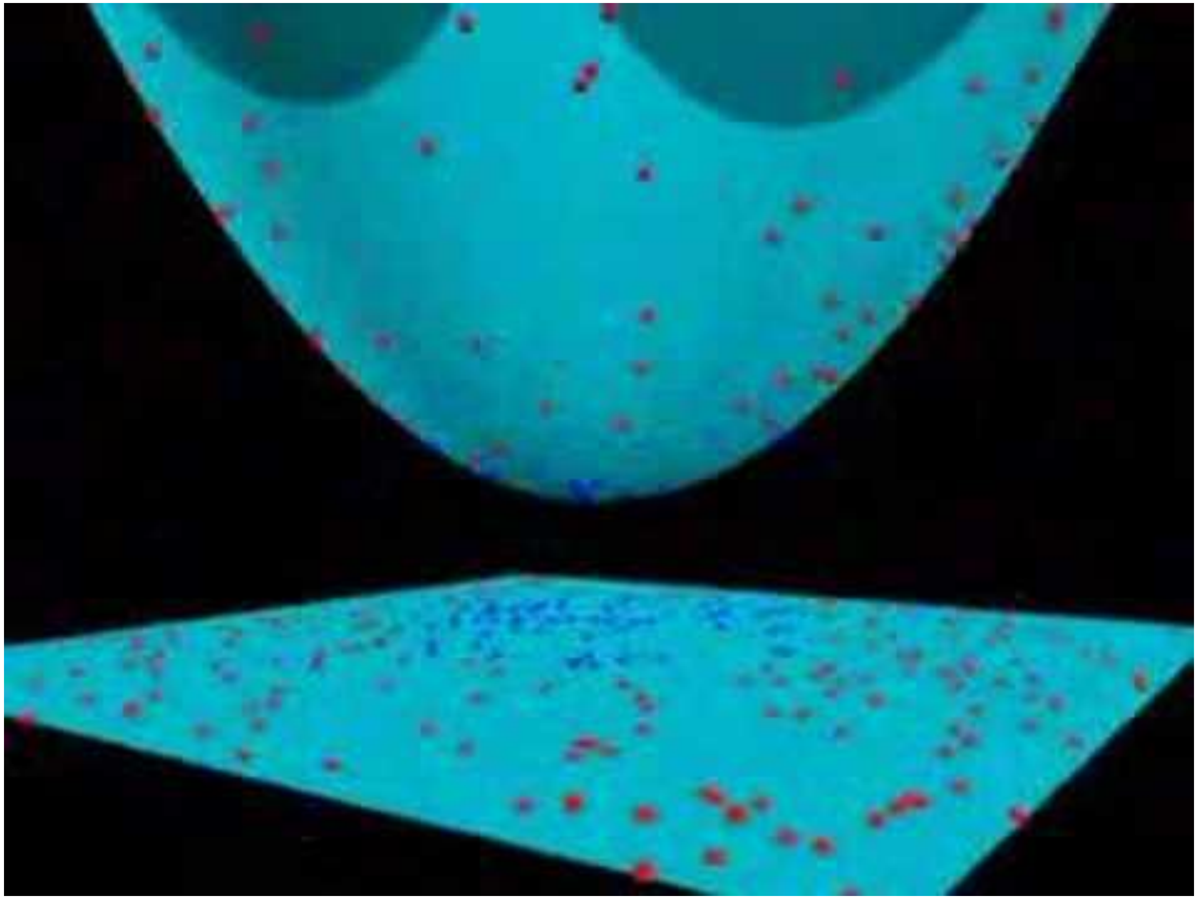


Image from: https://www.quora.com/What-is-a-Support-Vector-Machine

Example of supervised ML algorithm for classification

Support Vector Machines (SVM)

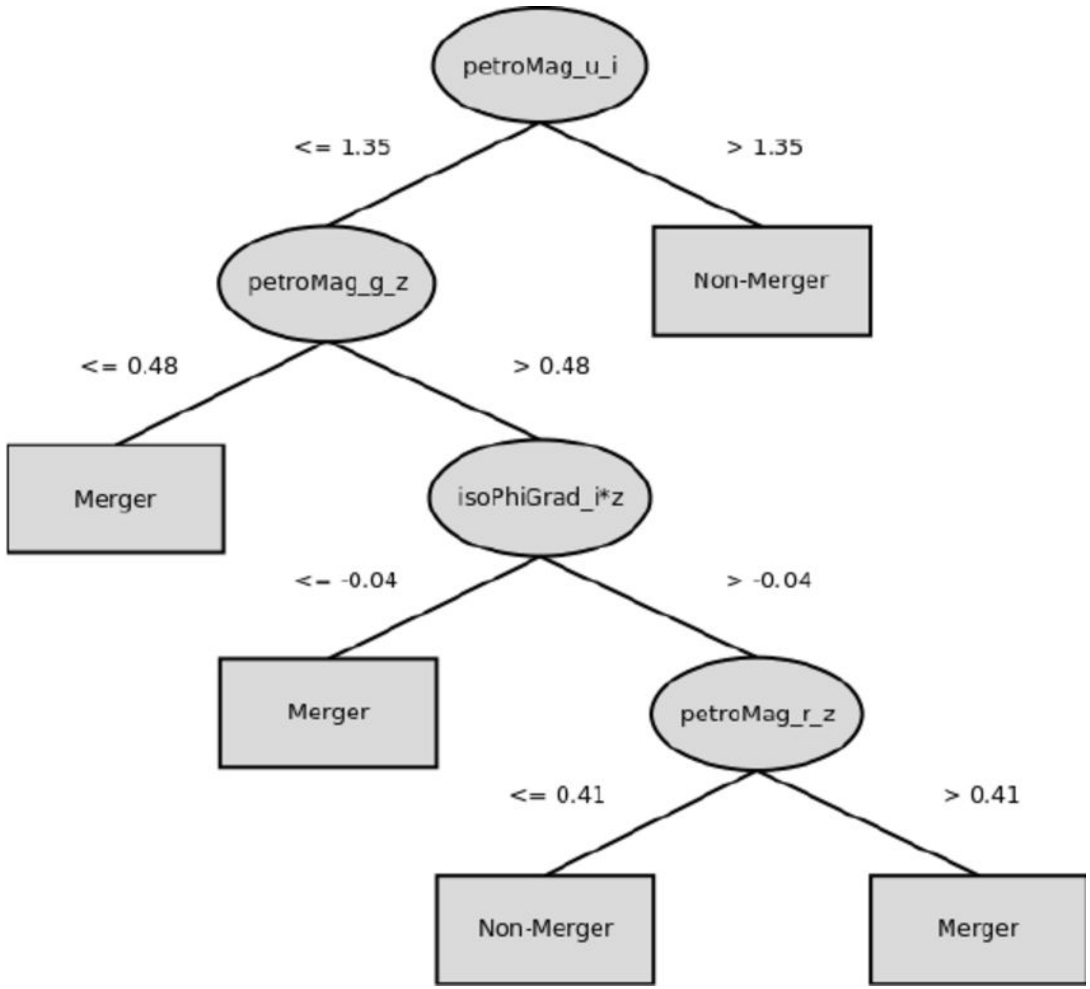
Search for hyperplanes - kernel trick



https://www.youtube.com/watch?v=3liCbRZPrZA

Example of supervised ML algorithm for classification

Decision Trees

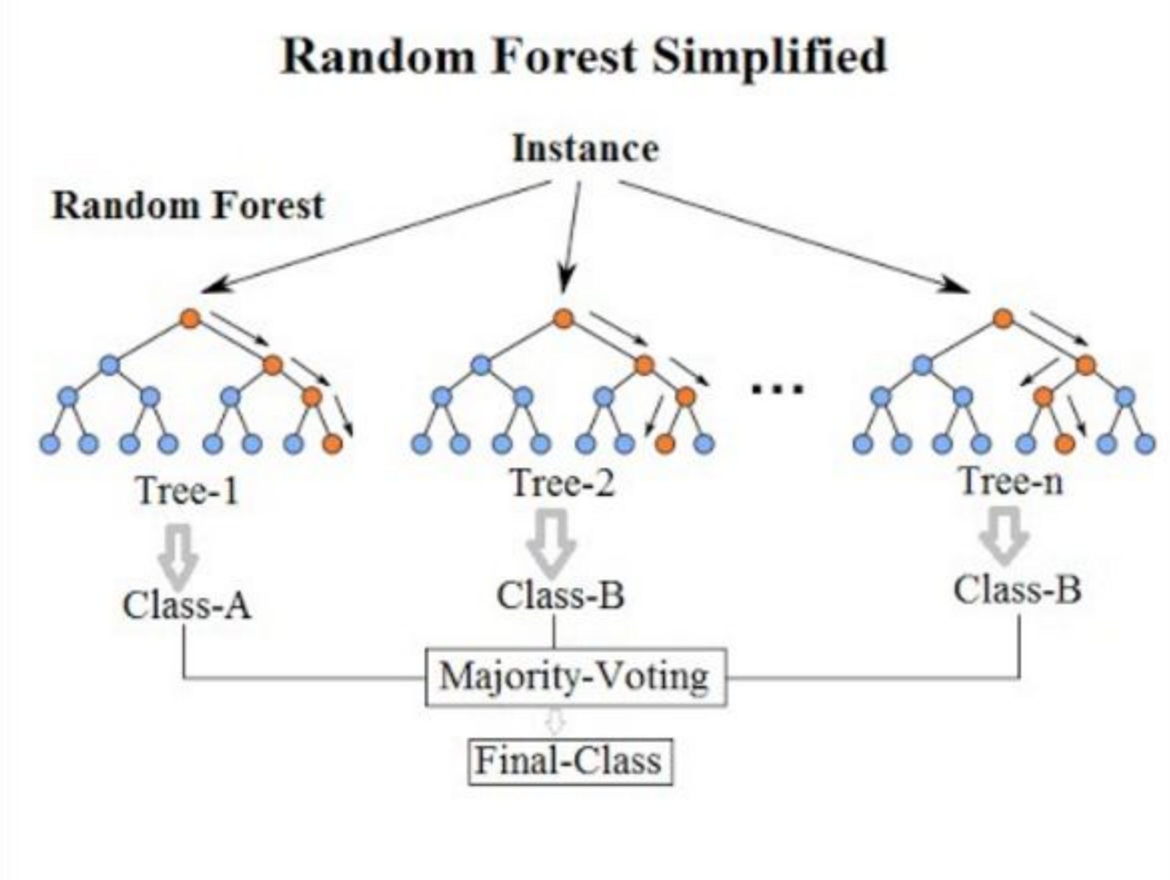


https://www.researchgate.net/figure/Visualization-of-decision-tree-built-using-the-weakest-mergers_fig4_221601336

Example of supervised ML algorithm for classification

Random Forests

Ensemble method

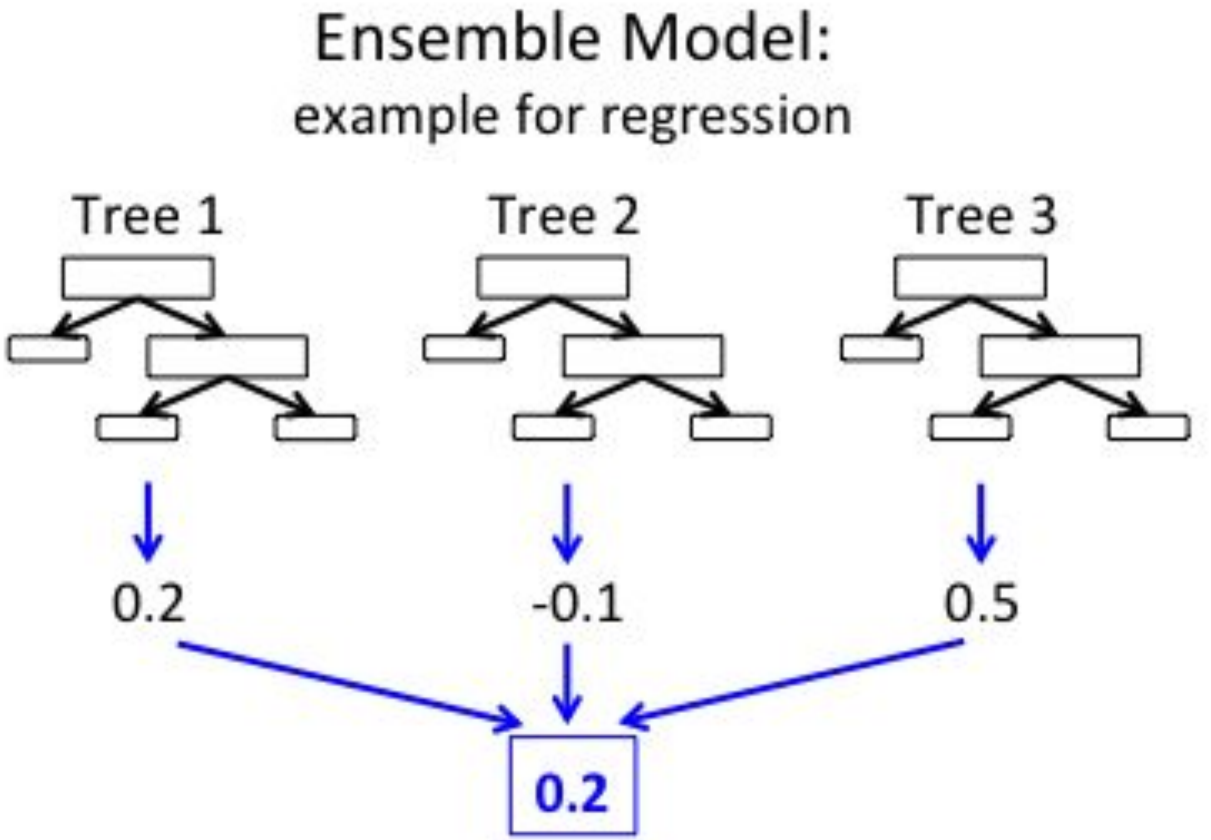


https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d

Example of supervised ML algorithm for regression

Random Forests

Ensemble method

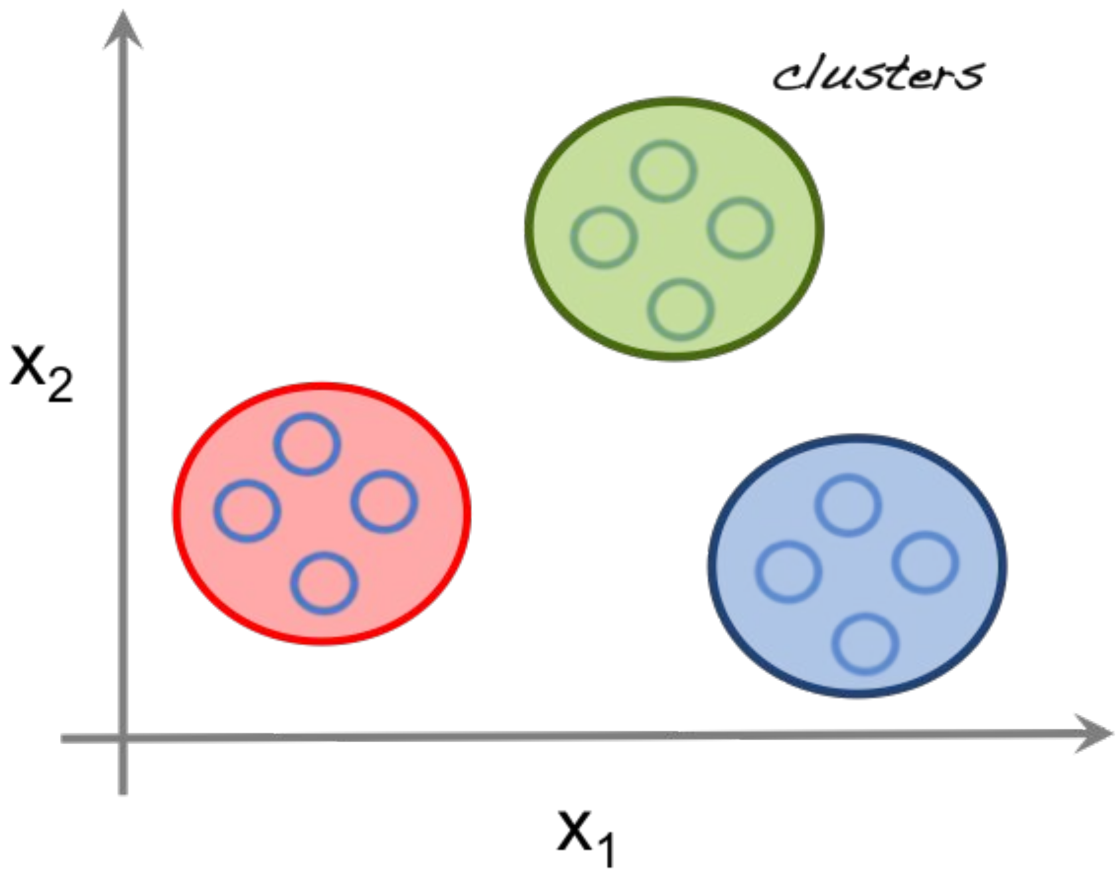


https://databricks.com/blog/2015/01/21/random-forests-and-boosting-in-mllib.html

Unsupervised Learning

Unsupervised Learning

Search for data structures



Unsupervised Learning

There is NO ground truth!

data	features
χ	set of all samples, x
~~Y~~	~~set of possible labels, y~~
$h_{\cancel{train}}$	learner: $y_{est;i} = h_{\cancel{train}}(x_i)$
~~L~~	~~Loss function~~

Data generation model:

$x_i \sim P_\chi$

~~f true labeling function, $y_i = f(x_i)$~~

~~$L_{data,f}(h) = P_{x\sim data}(h_{train}(x) \neq f(x))$~~

Unsupervised Learning

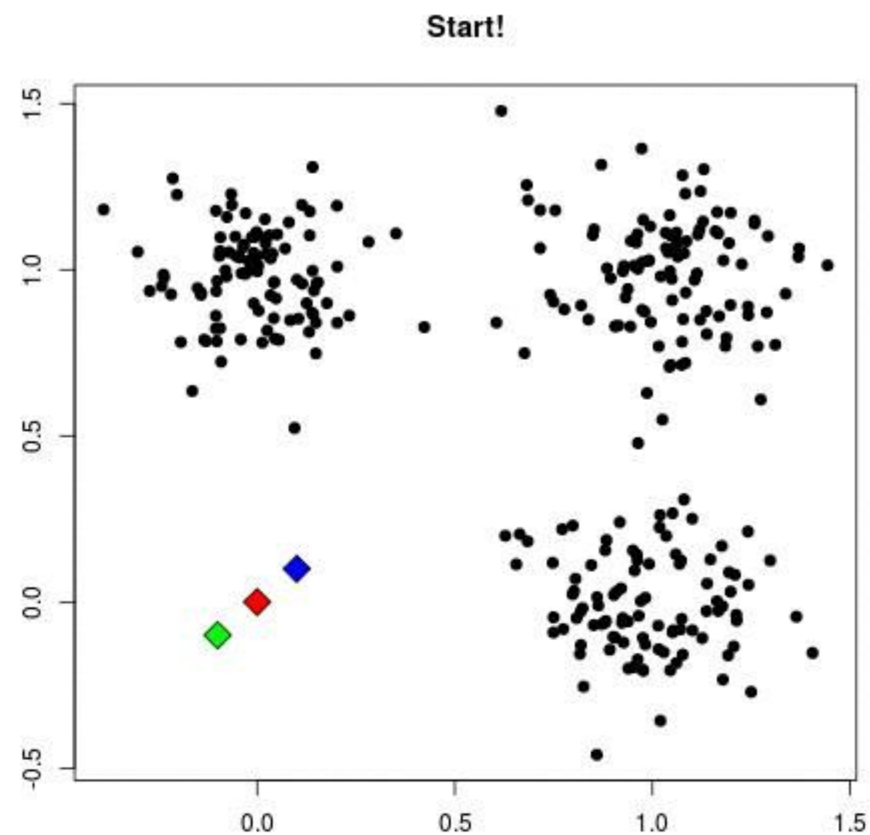
data	features
χ	set of all samples, x
h	learner: $y_{est;i} = h(x_i)$
$P(\chi)$	joint probability density

Goal: *characterize P*

Example of unsupervised ML algorithm

k-Means

Clustering



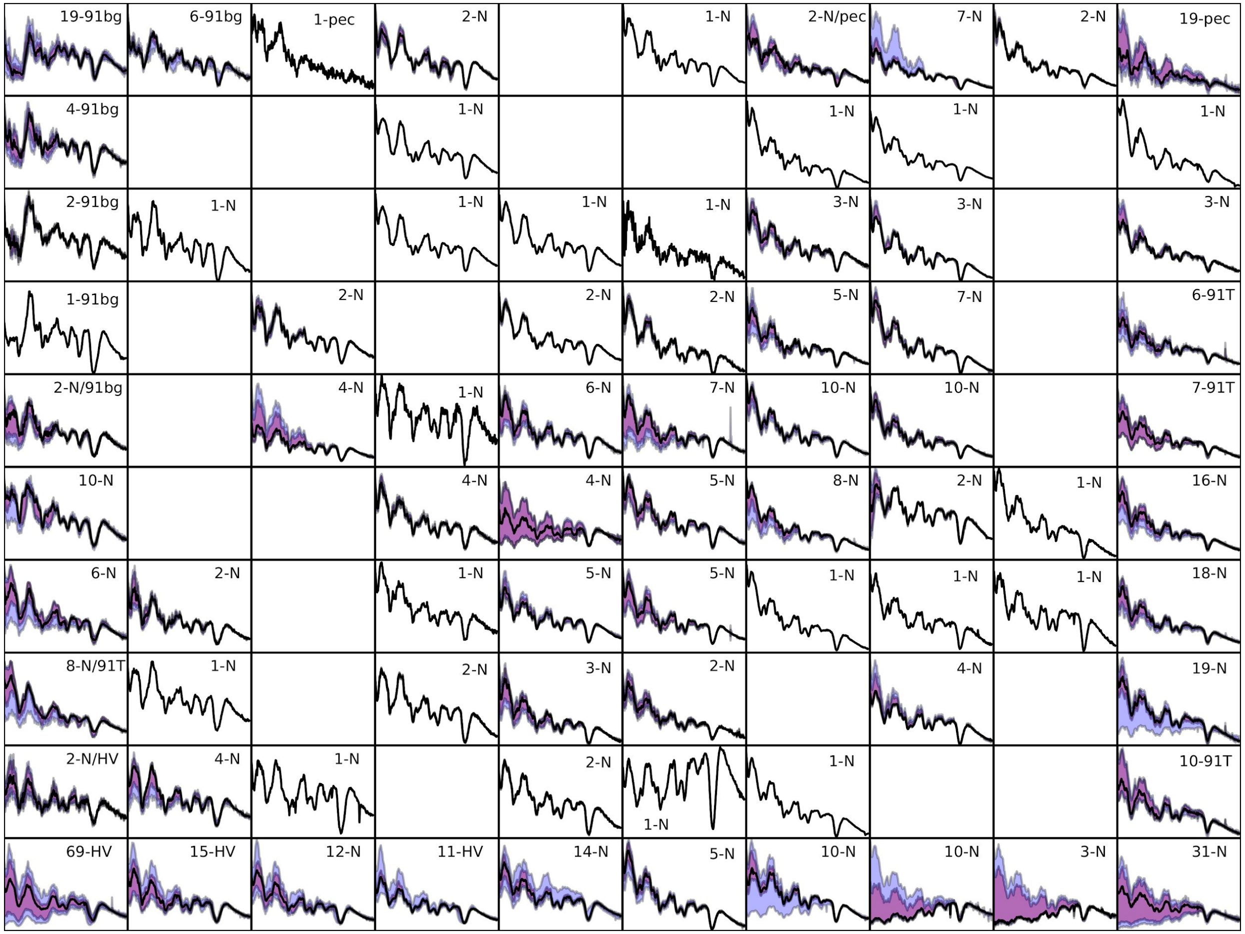
https://mubaris.com/2017/10/01/kmeans-clustering-in-python/

Example of unsupervised ML algorithm

Self-Organized Maps (SOM)

Clustering

Similar objects end up closer to each other



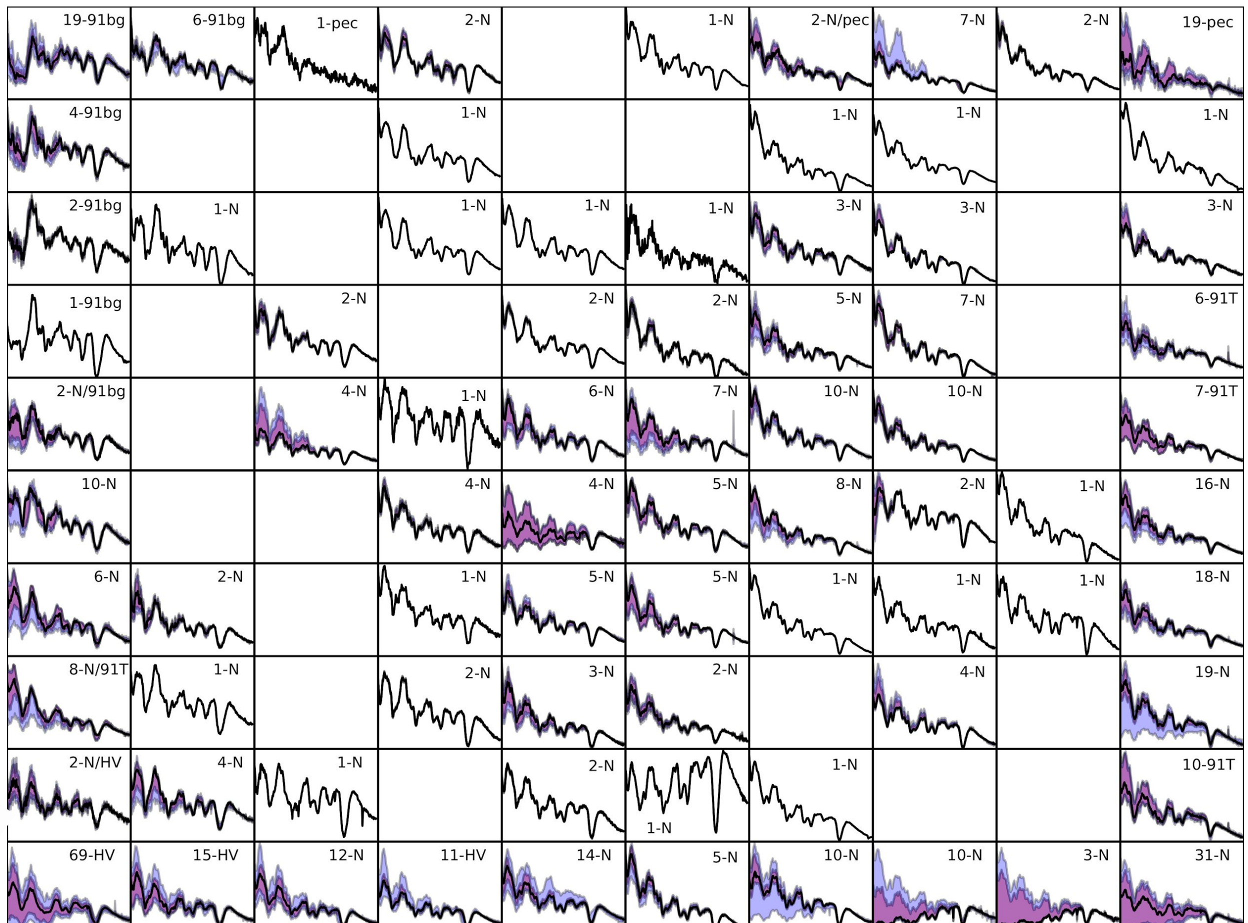
*Sasdelli, **Ishida** et al, 2016, MNRAS, 461, Issue 2, p.2044 - from CRP #2*

Example of unsupervised ML algorithm

Self-Organized Maps (SOM)

Clustering

Similar objects end up closer to each other



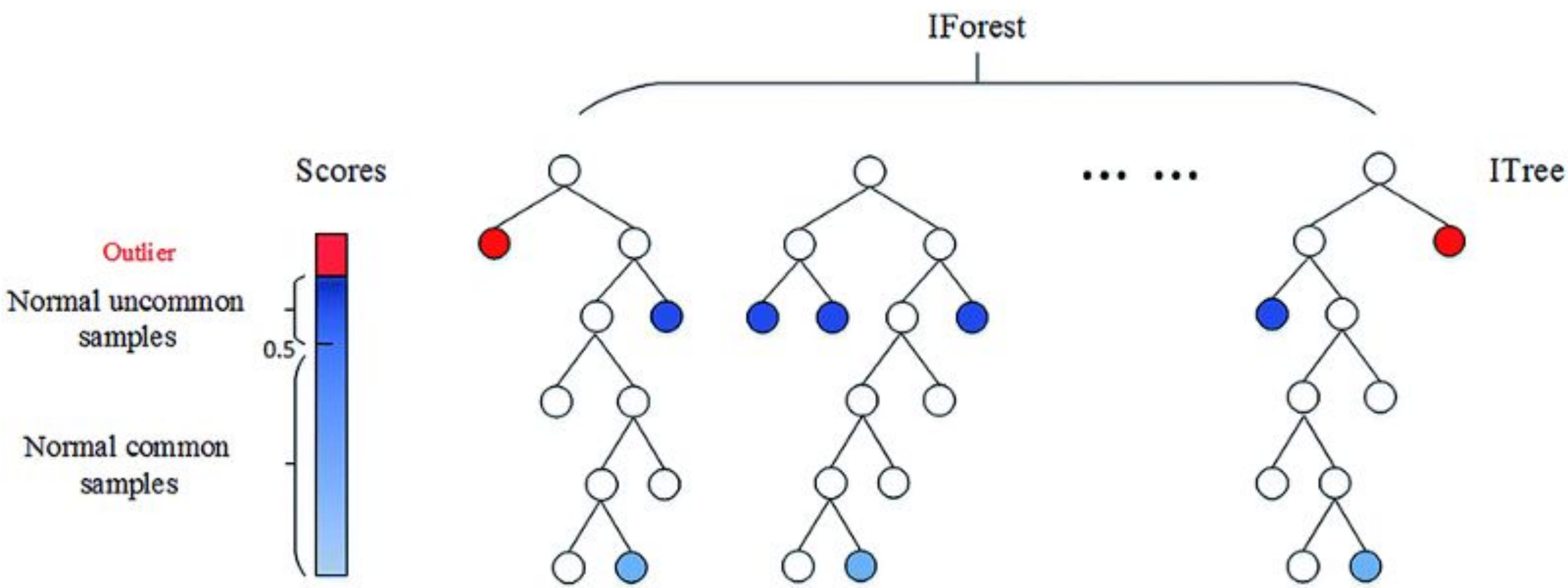
Are dimensionality reduction methods unsupervised learning?

Sasdelli, **Ishida** et al, 2016, MNRAS, 461, Issue 2, p.2044 - from CRP #2

Example of unsupervised ML algorithm

Isolation Forests

Outlier detection



https://towardsdatascience.com/a-brief-overview-of-outlier-detection-techniques-1e0b2c19e561

Summary

What you should remember from this session

1. The definition of learning is different for machines

Summary

What you should remember from this session

1. The definition of learning is different for machines

2. Supervised Machine Learning $\Longrightarrow$ Minimize loss function

Summary

What you should remember from this session

1. The definition of learning is different for machines

2. Supervised Machine Learning $\Longrightarrow$ Minimize loss function

3. Minimize loss function

$\Downarrow$

learn about the data generative model

Summary

What you should remember from this session

1. The definition of learning is different for machines

2. Supervised Machine Learning $\Longrightarrow$ Minimize loss function

3. Minimize loss function

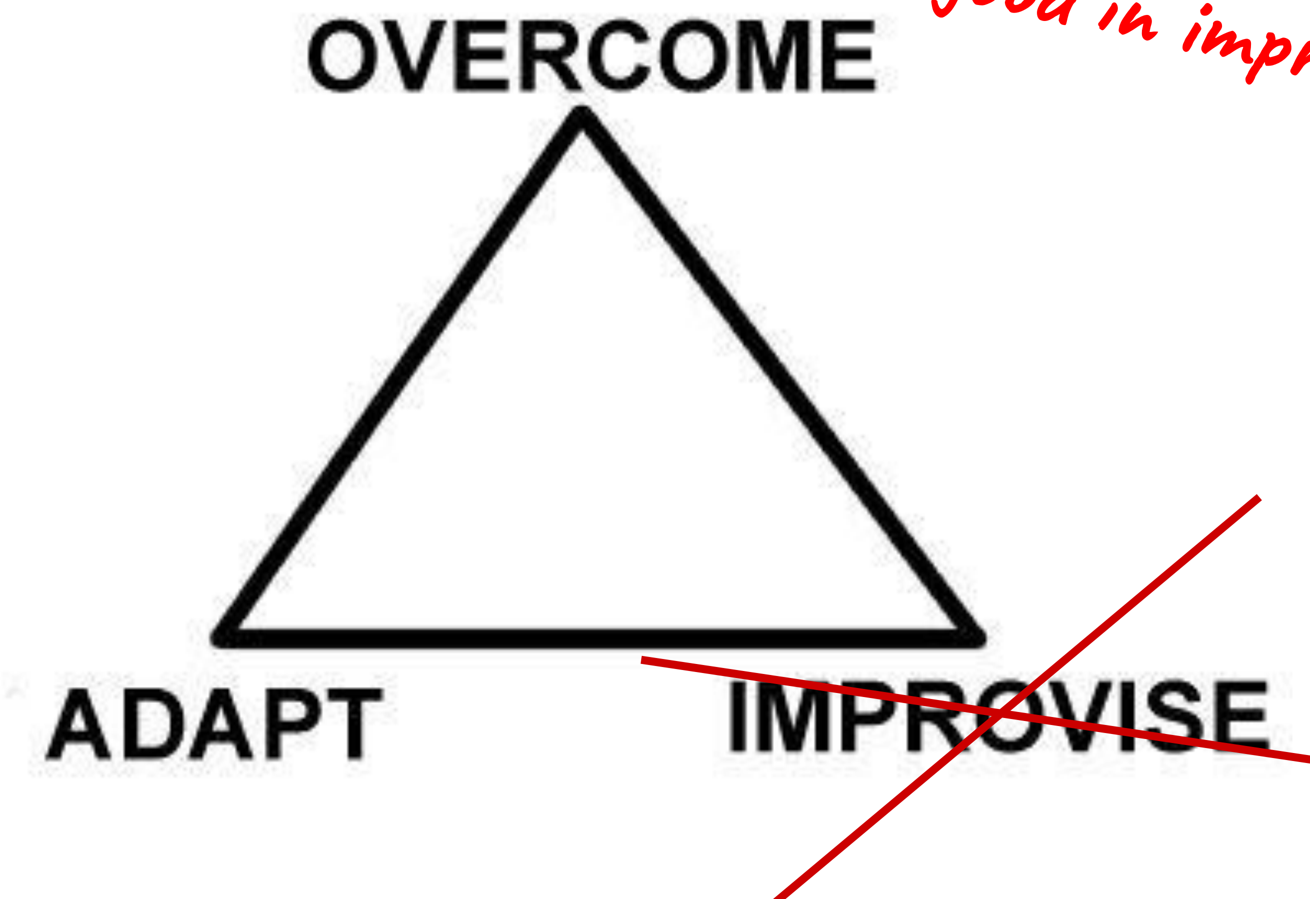
⇩

learn about the data generative model

4. The Machine Learning <u>algorithm</u>, or learner, is only 1 element in a larger Machine Learning <u>model</u> $\Longrightarrow$ you should pay attention to the other elements as well!

Learning *or the Power to Adapt*

Machines are not good in improvising!




Thank
You