3rd ASTERICS-OBELICS International School

8-12 April 2019, Annecy, France.

H2020-Astronomy ESFRI and Research Infrastructure Cluster
(Grant Agreement number: 653477).

Machine Learning Tutorial

I - Introduction

3rd ASTERICS-OBELICS International School
8-12 April 2019, Annecy - France

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

and

Alexandre Boucaud
Laboratoire AstroParticule & Cosmology (APC)
Paris, France

Clermont Ferrand, France

Truffade

COIN
Cosmostatistics Initiative

UNIVERSITÉ Clermont Auvergne

LSST Large Synoptic Survey Telescope

eZTF

Puy-de-Dome

Summary

I. Introduction to Machine Learning

II. Introduction to Neural Networks by *Alexandre Boucaud*

III. Deep Learning by *Alexandre Boucaud*

IV: Beyond textbook Machine Learning

Acknowledgments

Adam Miller

Andrew Ng

Fabian Gieseke

David Hogg

David Kirby

Daniela Huppenkothen

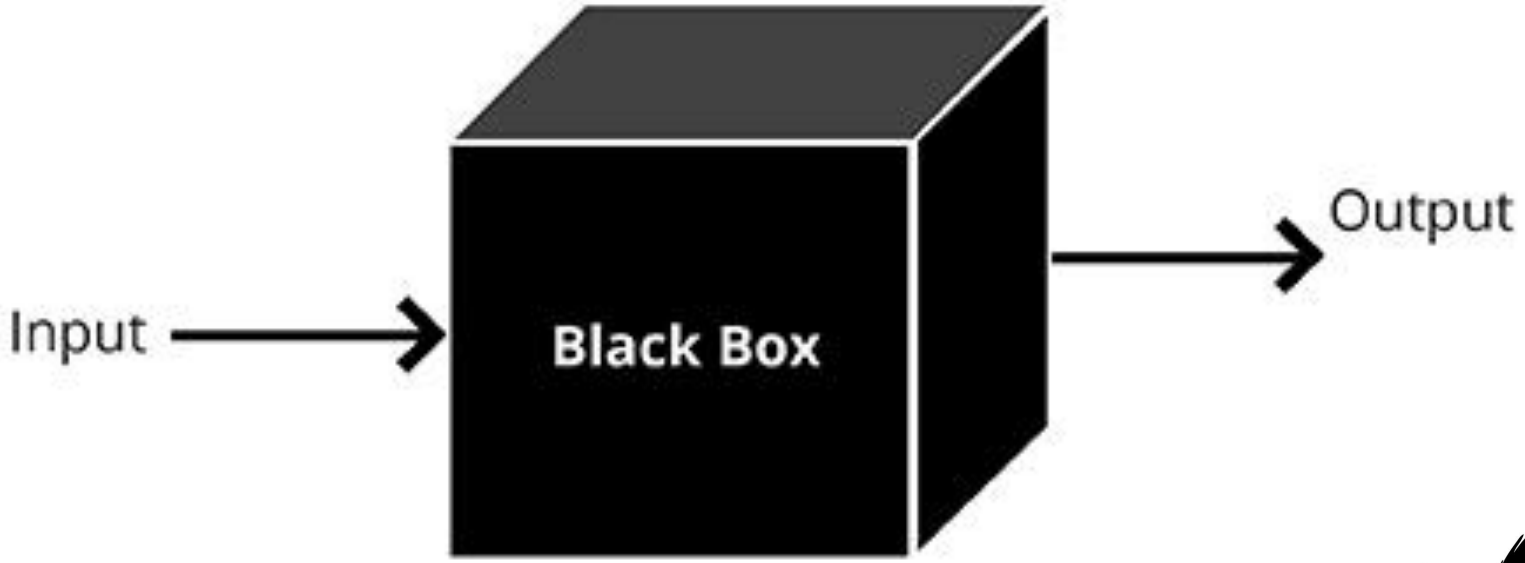
Jake VanderPlas

Joshua Bloom

Ricardo Vilalta

The Cosmostatistics Initiative

Beware of Black Boxes!!!



Disclaimer 2

for now …

What is learning ?

(real question, google it!)

What is learning ?

"A relatively permanent change in behaviour due to past experiences."

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

What is Machine Learning ?

(you know what to do)

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

Example 1: the Iris dataset

Presented by R. A. Fisher (1936)

50 samples of each class

Petal length	Petal width	Sepal length	Sepal width	Class
5.1	3.5	1.4	0.2	Setosa
6.5	1.0	4.5	1.3	Versicolor
5.9	3.0	5.1	1.8	Virginica
...	...	...	...	...

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

Petal length	Petal width	Sepal length	Sepal width	Class
5.1	3.5	1.4	0.2	Setosa
6.5	1.0	4.5	1.3	Versicolor
5.9	3.0	5.1	1.8	Virginica
...	...	...	...	...

Setosa

Versicolor

Virginica

What can you hope to learn from this?

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

Petal length	Petal width	Sepal length	Sepal width	Class
5.1	3.5	1.4	0.2	Setosa
6.5	1.0	4.5	1.3	Versicolor
5.9	3.0	5.1	1.8	Virginica
...	...	...	...	...

experience E → data

task T → classification

performance P → accuracy

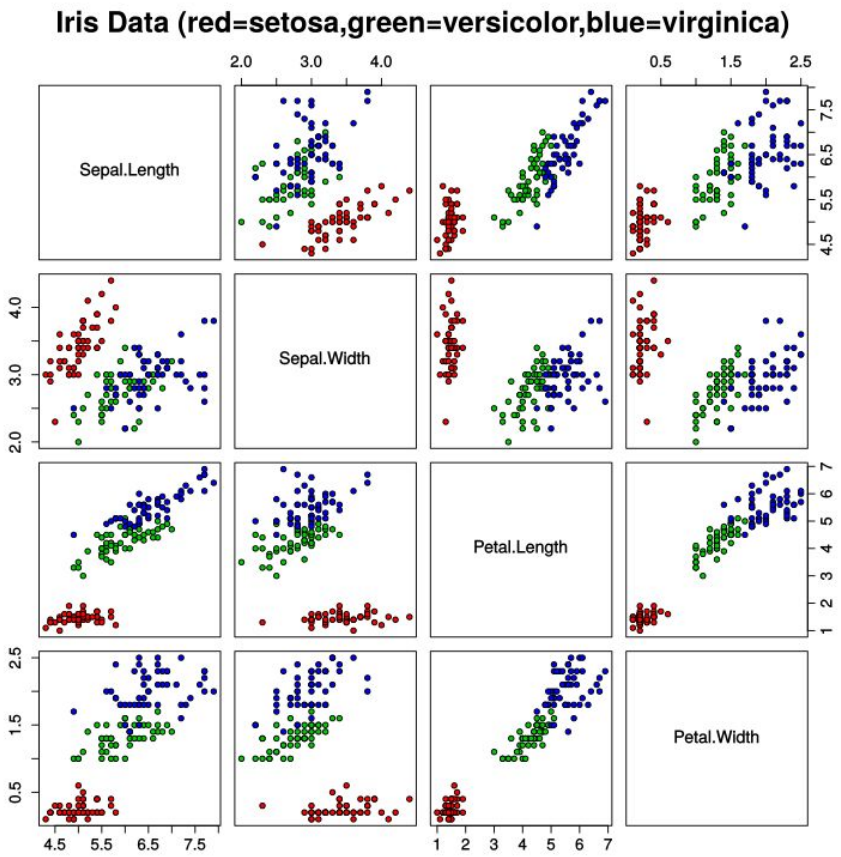
program L → strategy

What are your expectations?

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 →

classification

performance P → accuracy

program L → strategy

What are your expectations?

Example 2: SDSS

Notebook 1

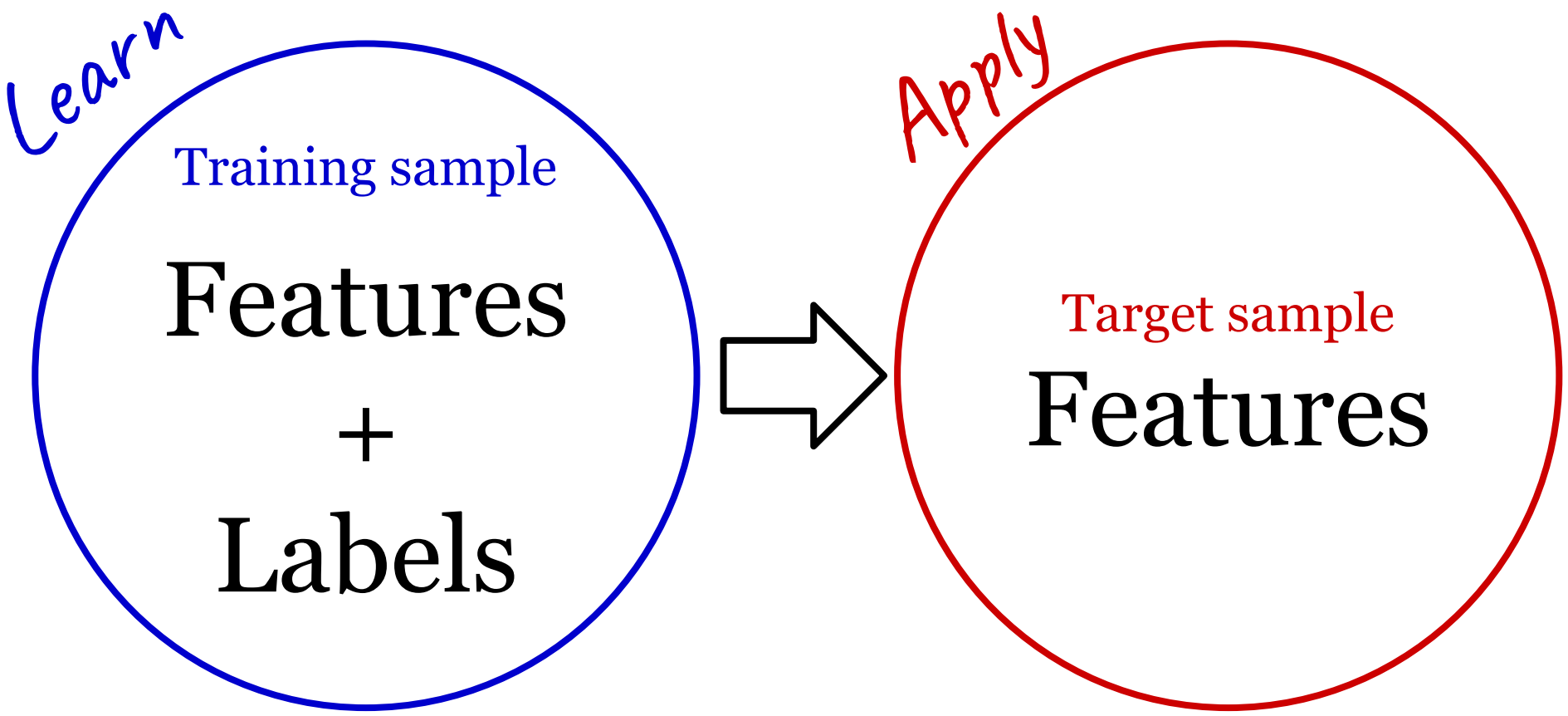
EDA_SDSS.ipynb

Supervised Learning

Categories of Machine Learning:

Supervised Learning

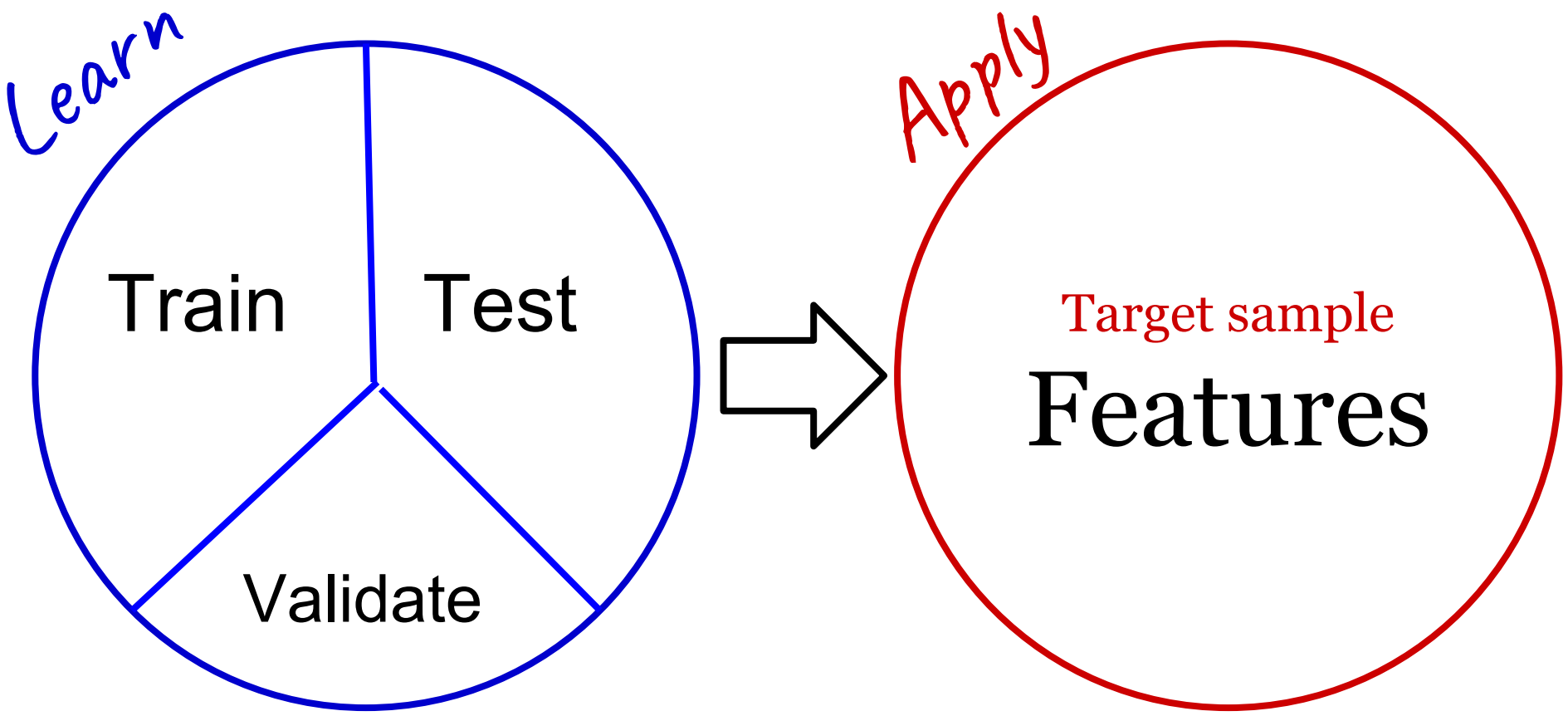
Learn by example



Categories of Machine Learning:

Supervised Learning

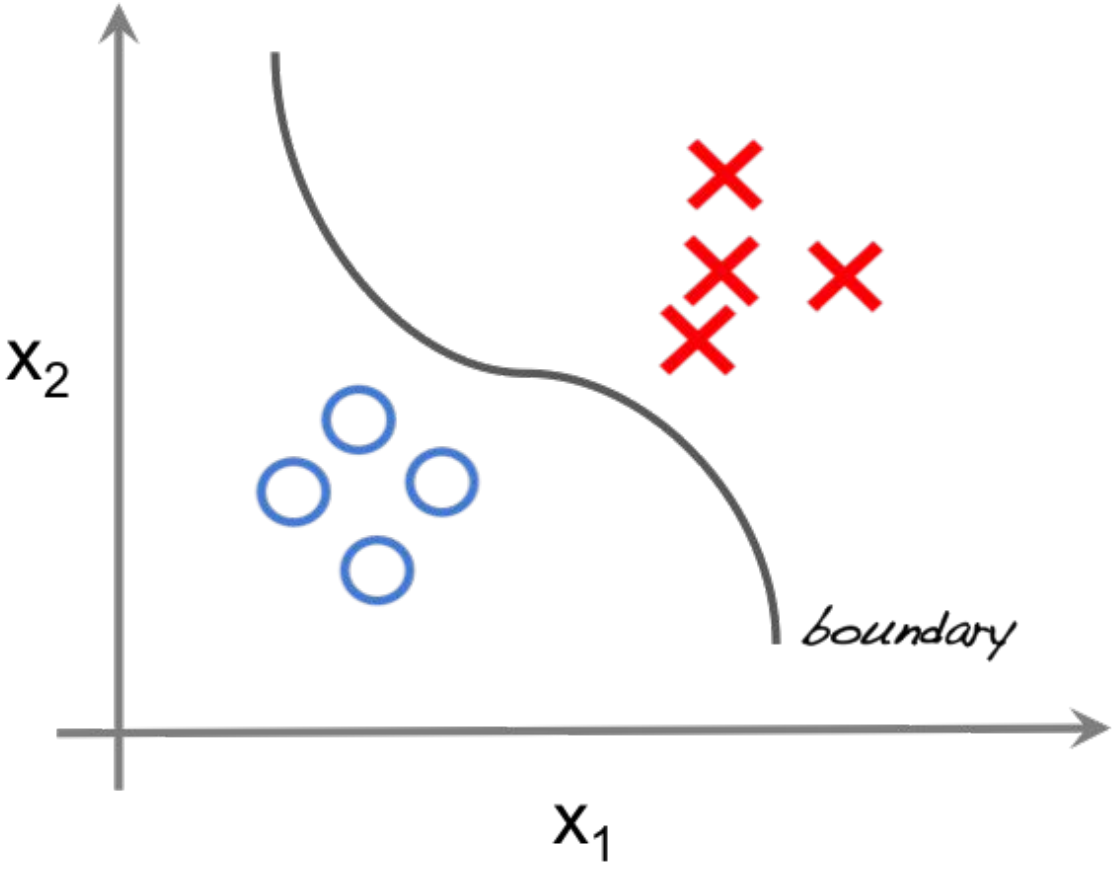
Learn by example



Categories of Machine Learning:

Supervised Learning

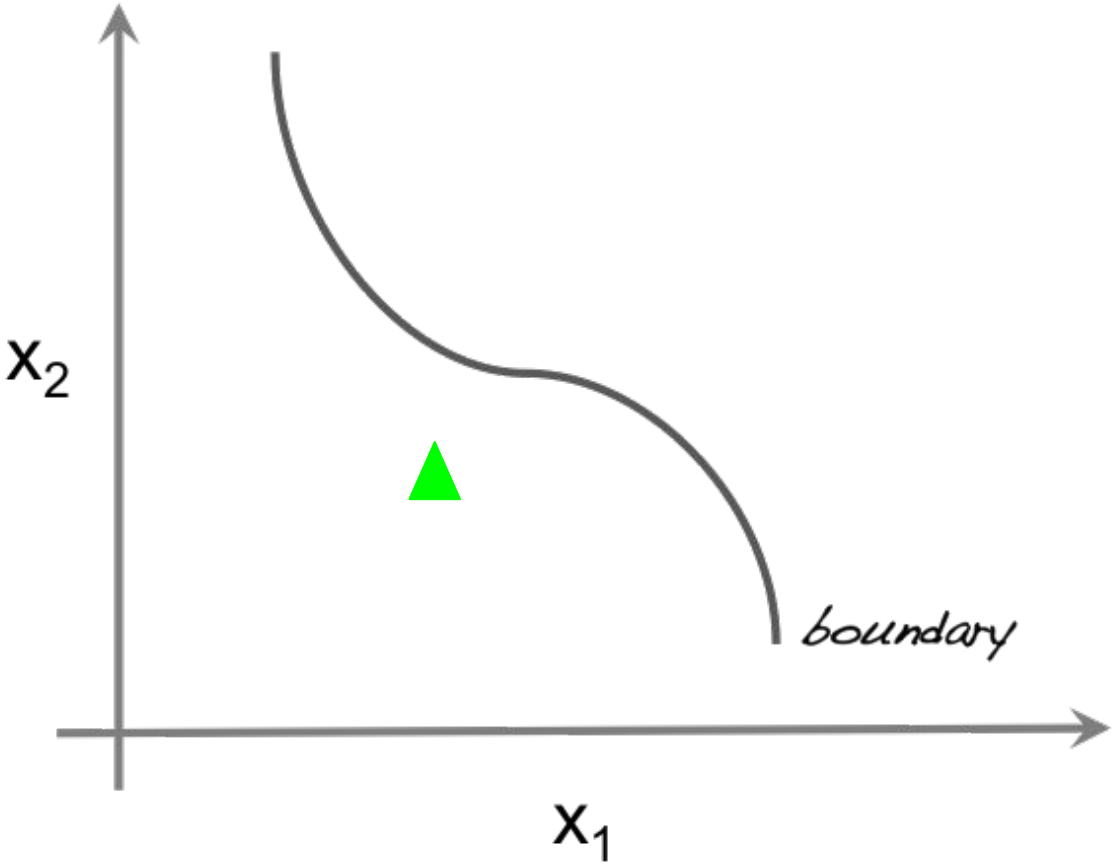
Classification



Categories of Machine Learning:

Supervised Learning

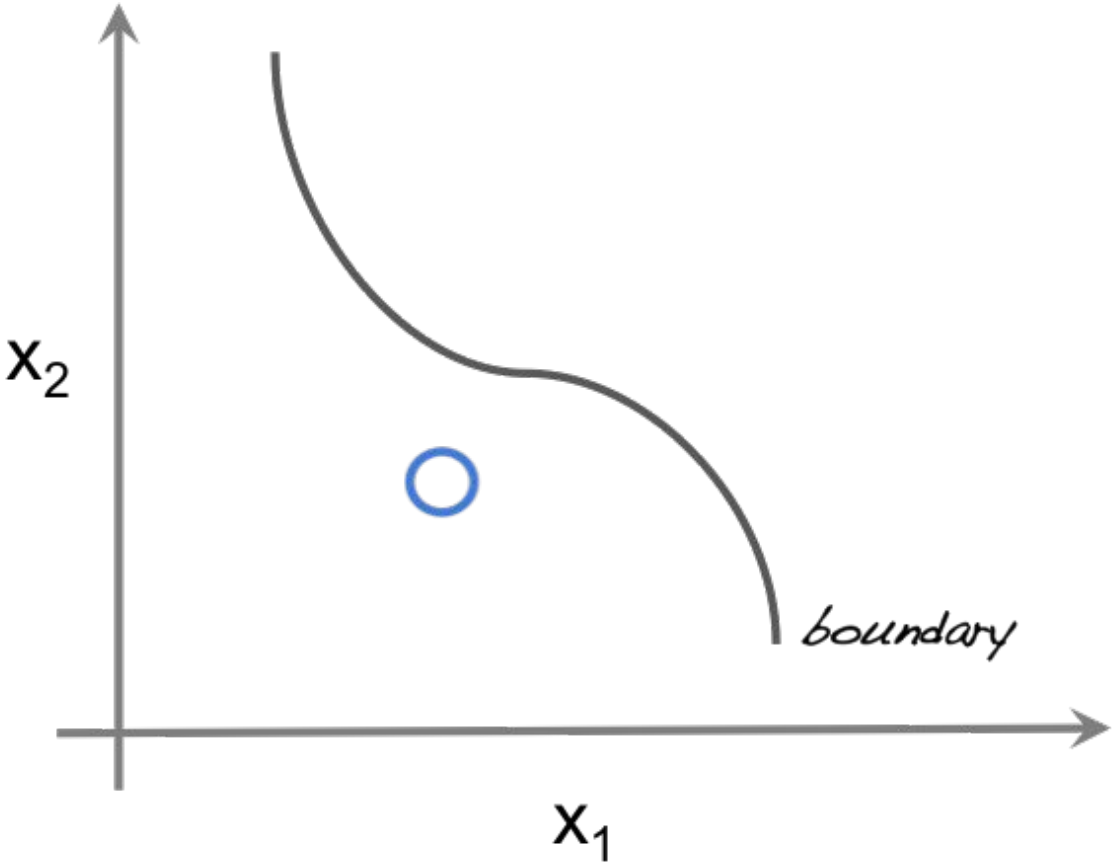
Classification



Categories of Machine Learning:

Supervised Learning

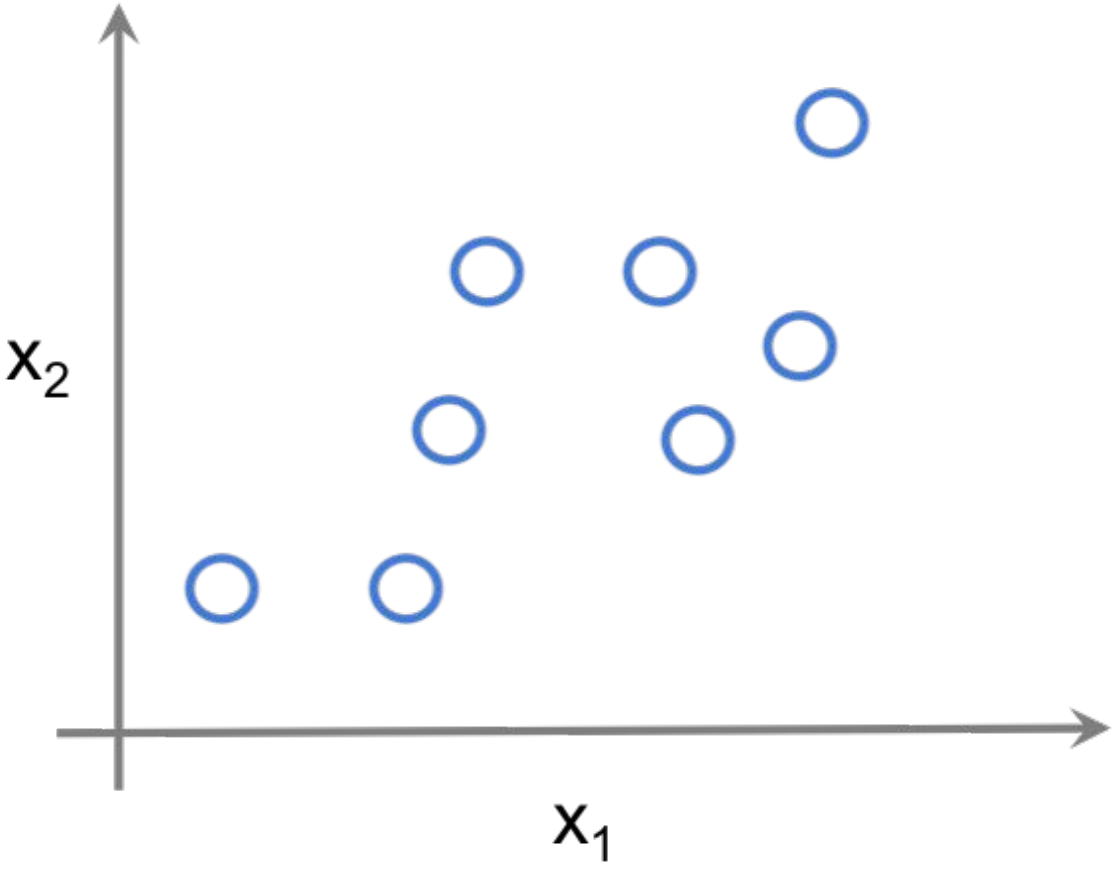
Classification



Categories of Machine Learning:

Supervised Learning

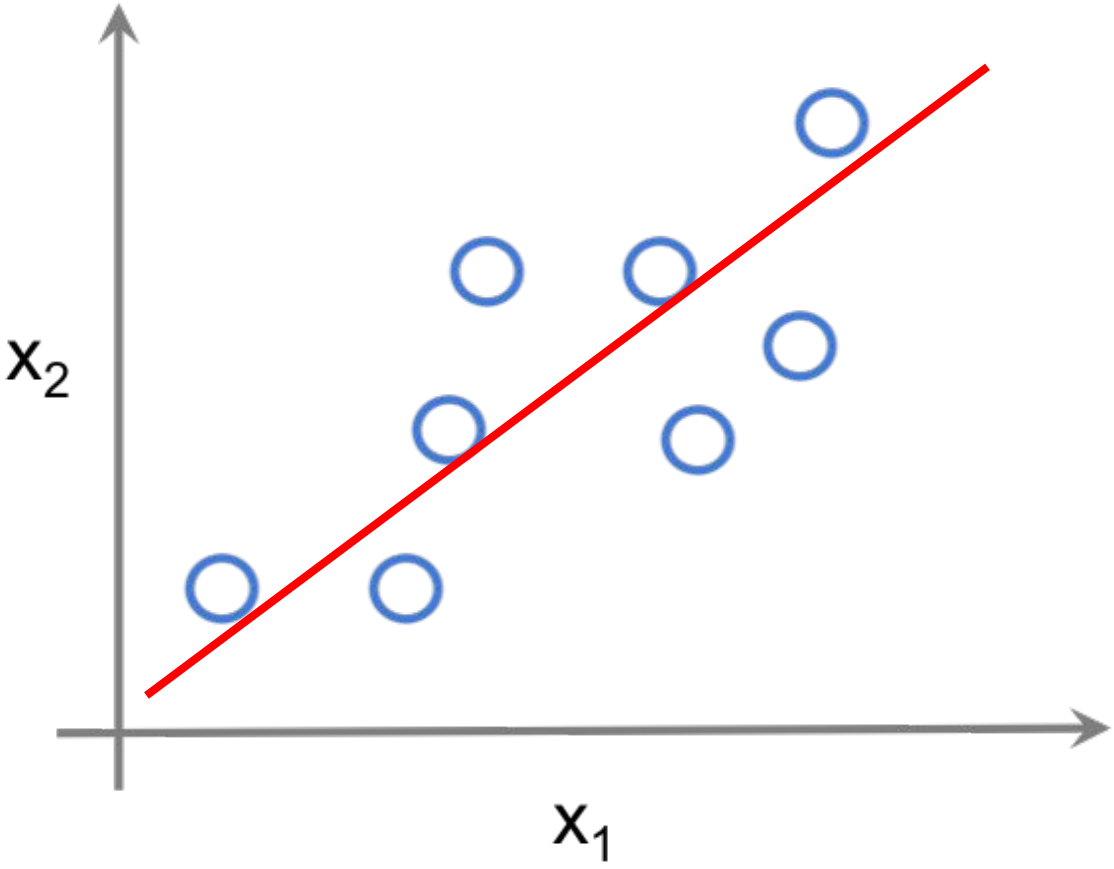
Regression



Categories of Machine Learning:

Supervised Learning

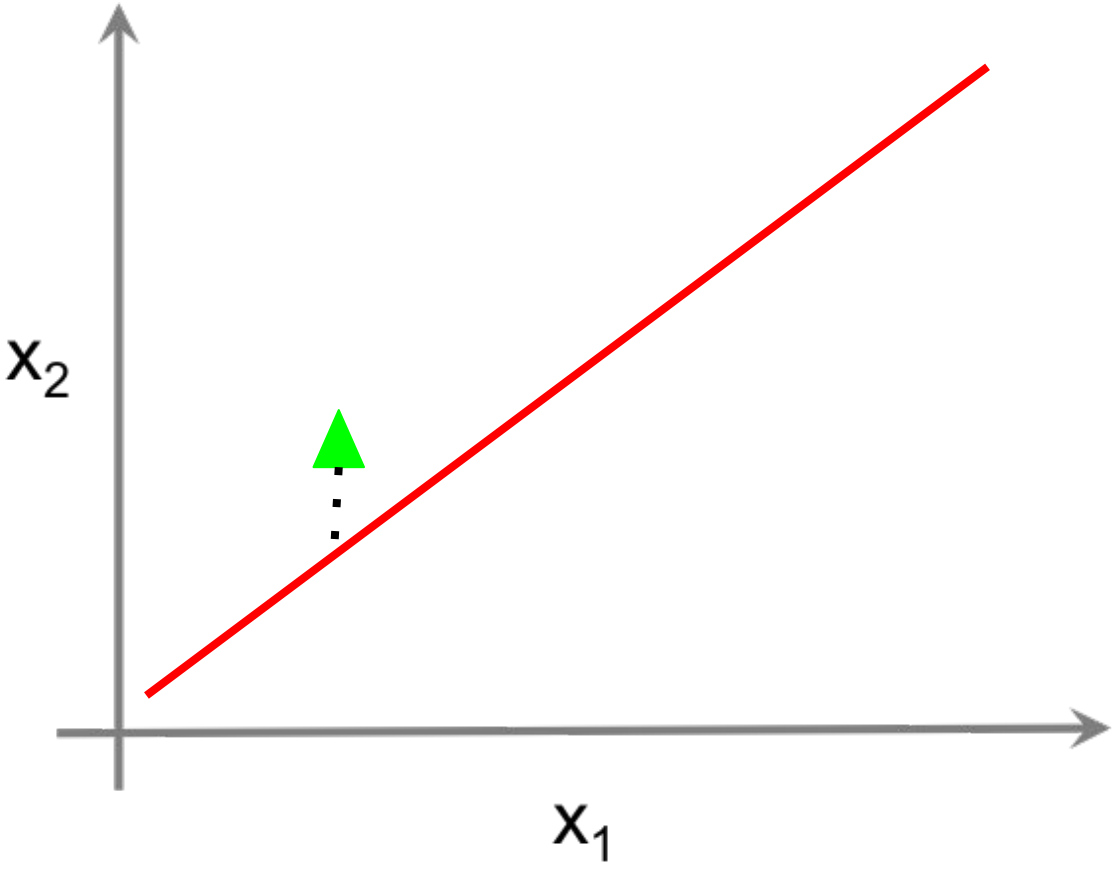
Regression



Categories of Machine Learning:

Supervised Learning

Regression



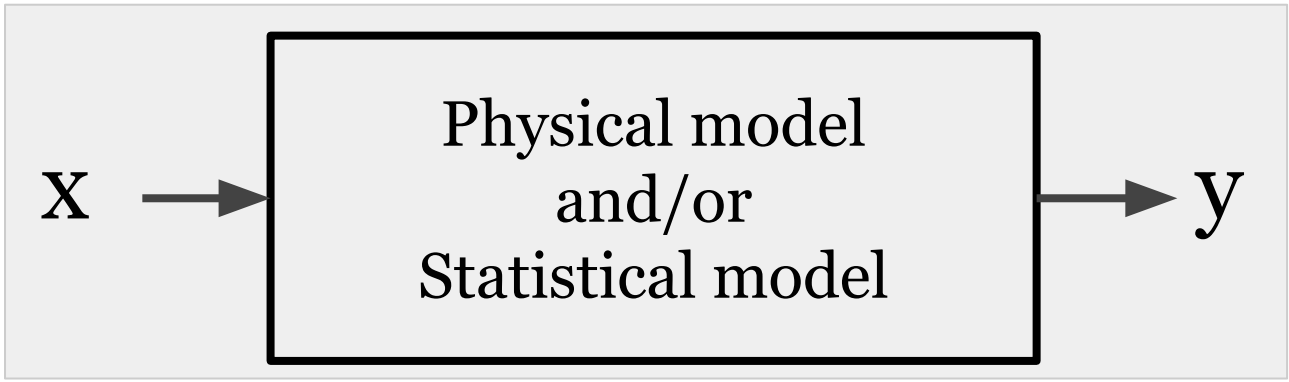
What is the difference between machine learning and a traditional data modelling?

Hypothesis:

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

Hypothesis:

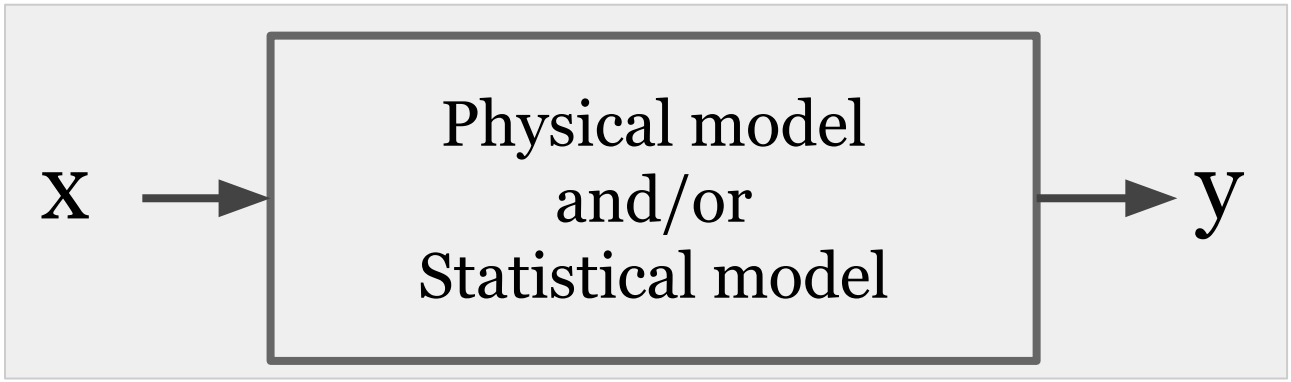
Traditional data modeling:



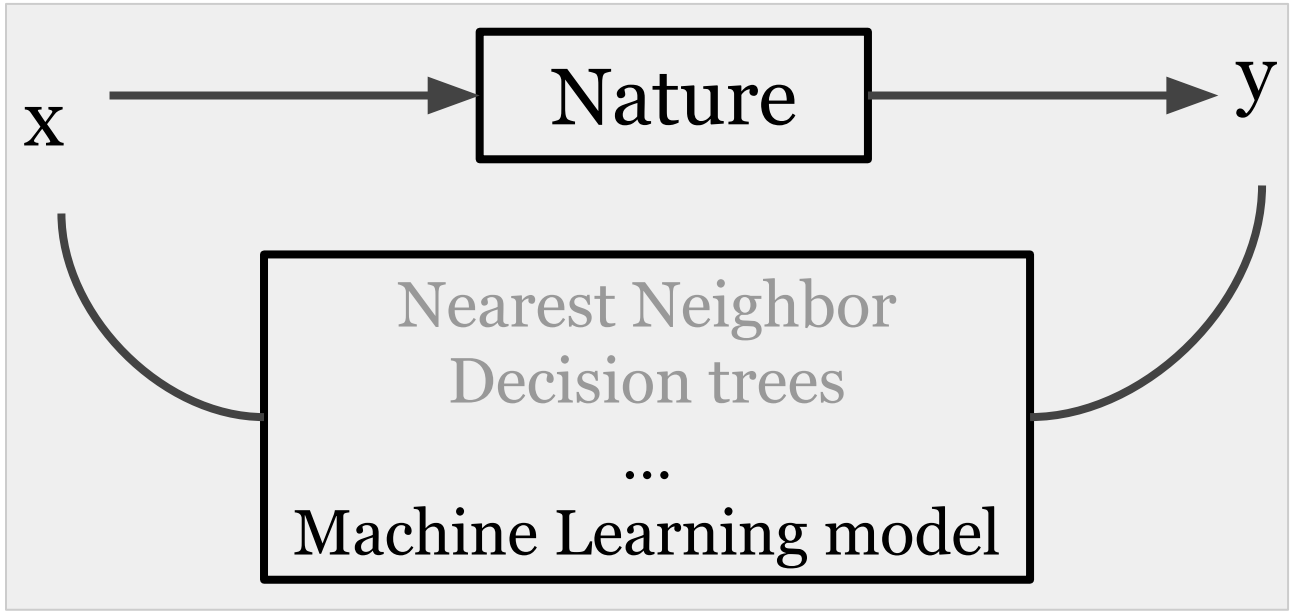
Breiman, L., *Statistical Modeling: The Two Cultures*, Stat. SciVolume 16 (2001)

Hypothesis:

Traditional data modeling:



Algorithmic modeling:



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

Data generation model:

$x_i \sim P_\chi$

$f \rightarrow$ 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

Machine Learning algorithm

X set of possible examples, x

Y set of possible labels, y

L Loss function

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

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

Machine Learning algorithm

1. Linear Regression
2. Logistic Regression
3. Decision Tree
4. SVM
5. Naive Bayes
6. kNN
7. K-Means
8. Random Forest
9. Dimensionality Reduction Algorithms
10. Gradient Boosting algorithms
    1. GBM
    2. XGBoost
    3. LightGBM
    4. CatBoost

https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/

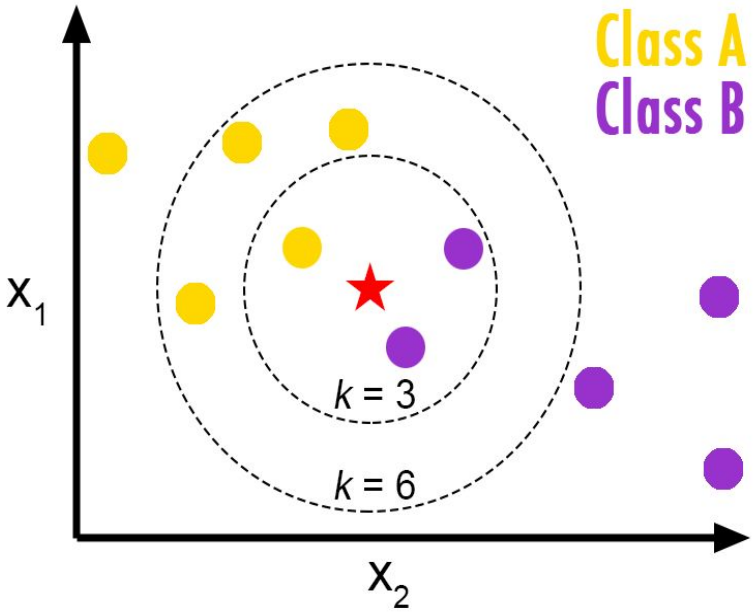
Example of supervised ML algorithm for classification

k-Nearest Neighbor (kNN)

Distance based

Hypothesis:

Objects from similar classes are clustered together



data → $\{x_1, x_2\}$

Possible labels → {A, B}

Learner → nearest neighbor

Loss-function → distance

Notebook 2

Classification.ipynb

Example of supervised ML algorithm for classification

Support Vector Machines (SVM)

Search for hyperplanes

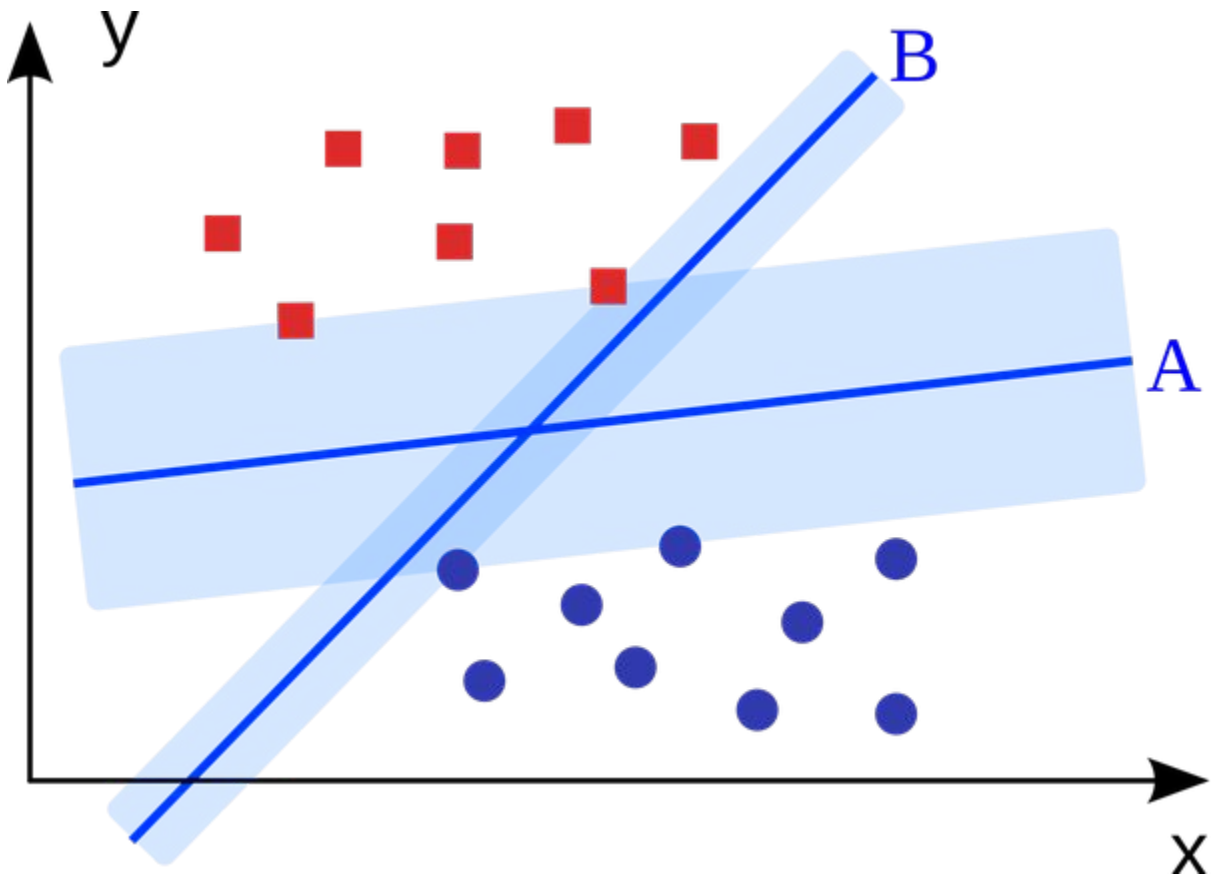
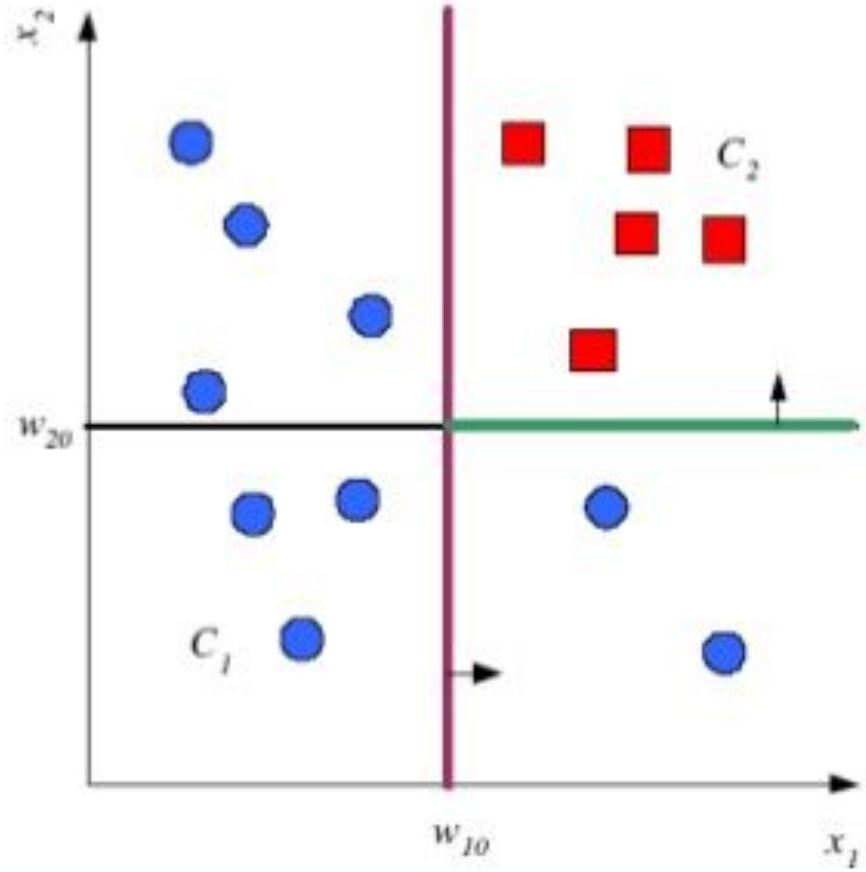
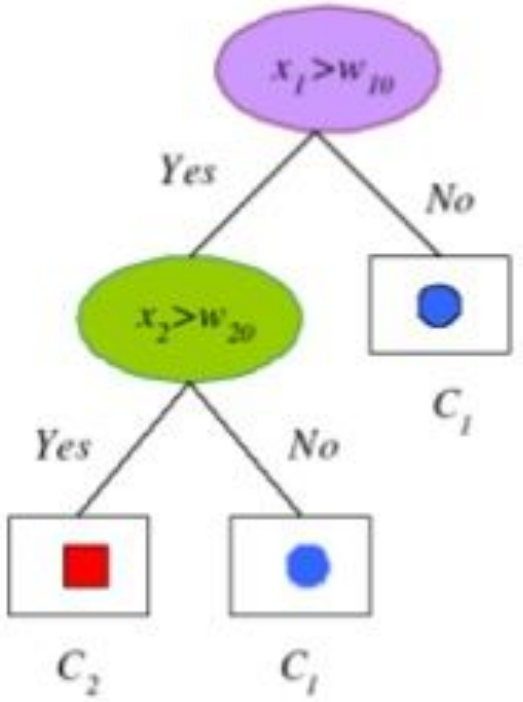


Image from: https://www.quora.com/What-is-a-Support-Vector-Machine

Example of supervised ML algorithm

Decision Trees



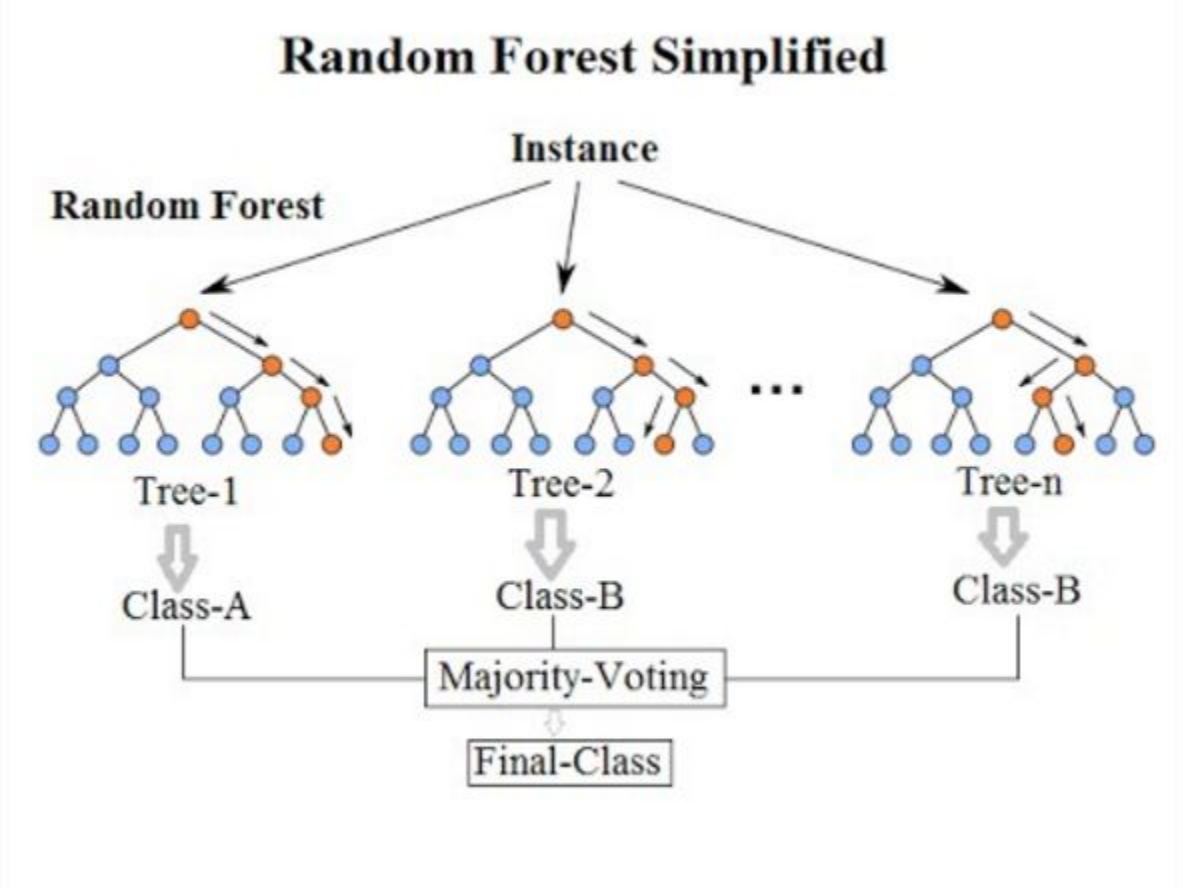


http://www.lewisgavin.co.uk/Machine-Learning-Decision-Tree/

Example of supervised ML algorithm

Random Forests

Ensemble method



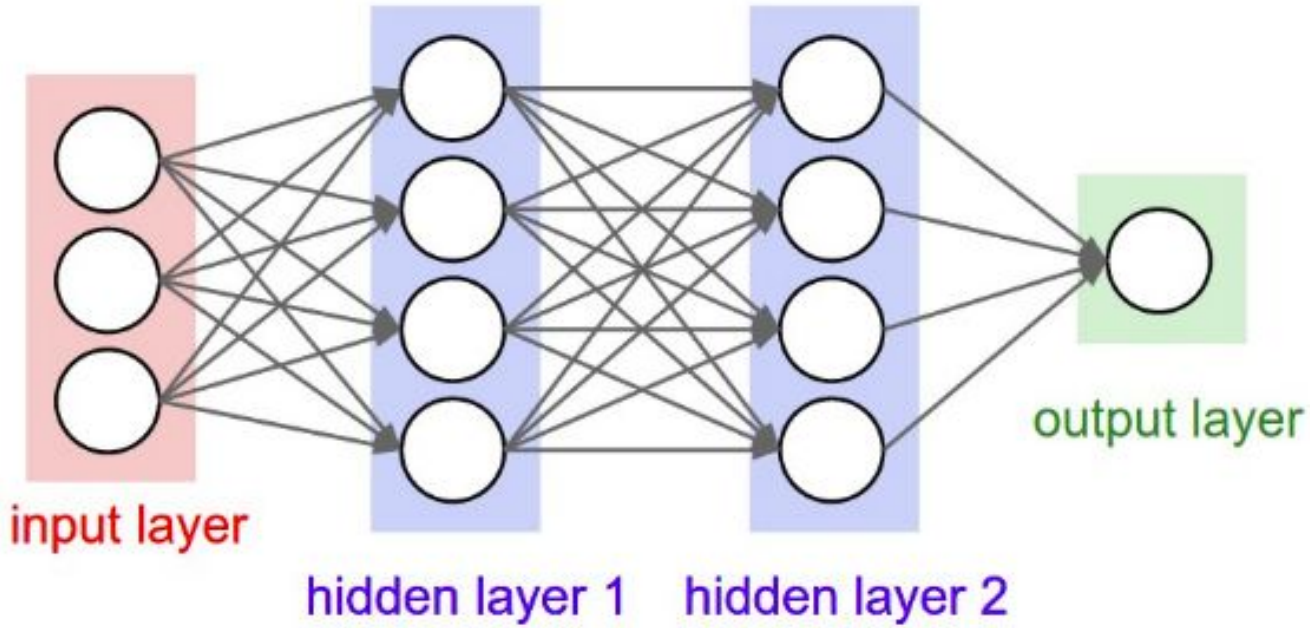
https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d

Example of supervised ML algorithm:

(Deep) Neural Network

All layers internal to the network (not input or output layer) are considered hidden layers.

See Alex's tutorial tomorrow



Slide by Alexandre Boucaud, ADA IX, 2018

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

Data generation model:

$x_i \sim P_\chi$

$f \rightarrow$ 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

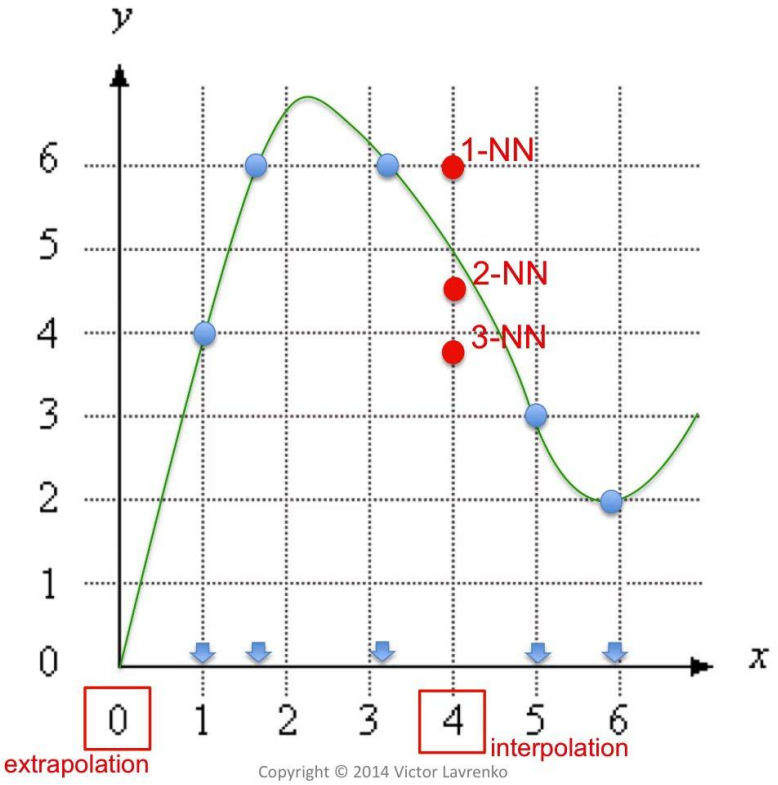
It also works for regression...

Example of supervised ML algorithm for regression

k-Nearest Neighbor (kNN)

Distance based

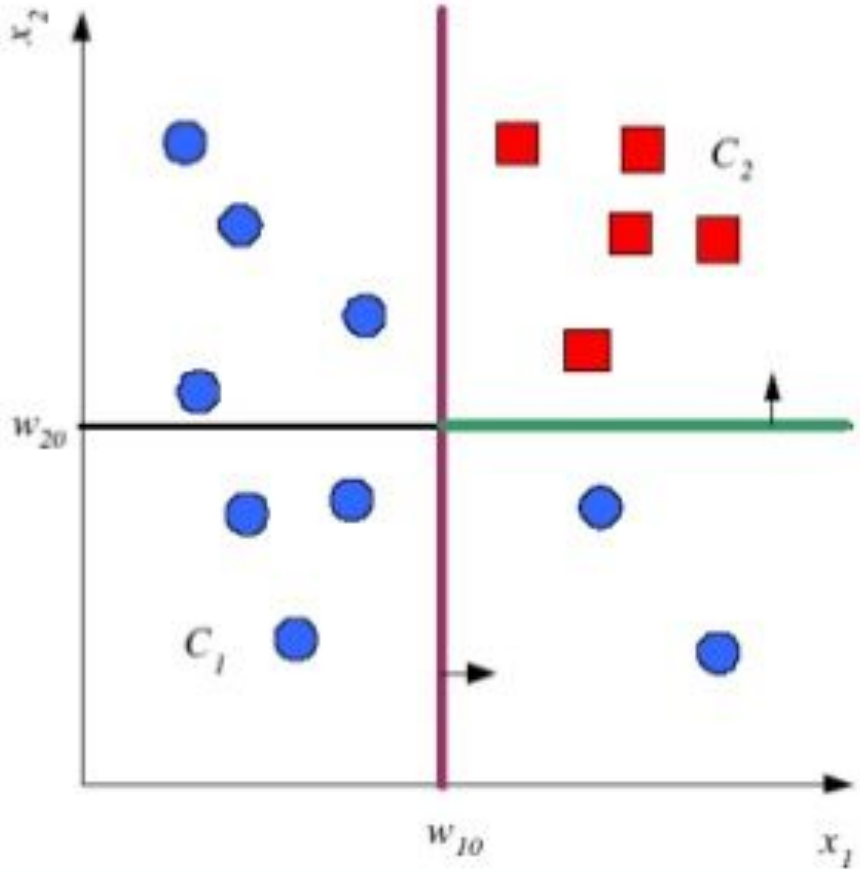
Example: kNN regression in 1-d

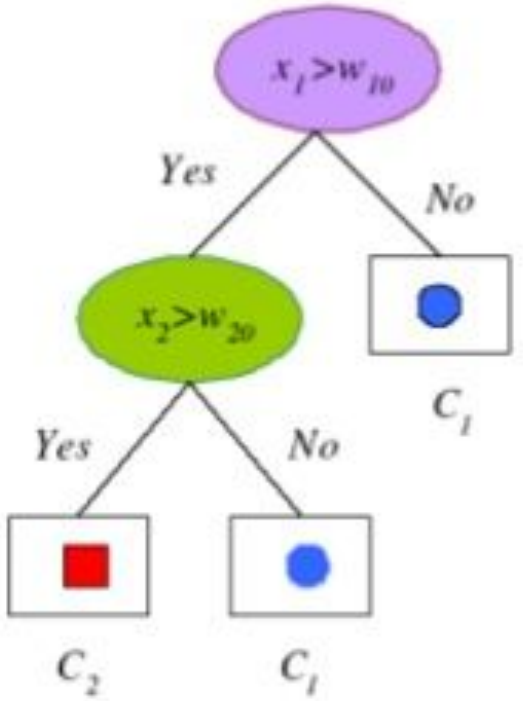


https://www.youtube.com/watch?v=3lp5CmSwrHI

Example of supervised ML algorithm for classification

Decision Trees



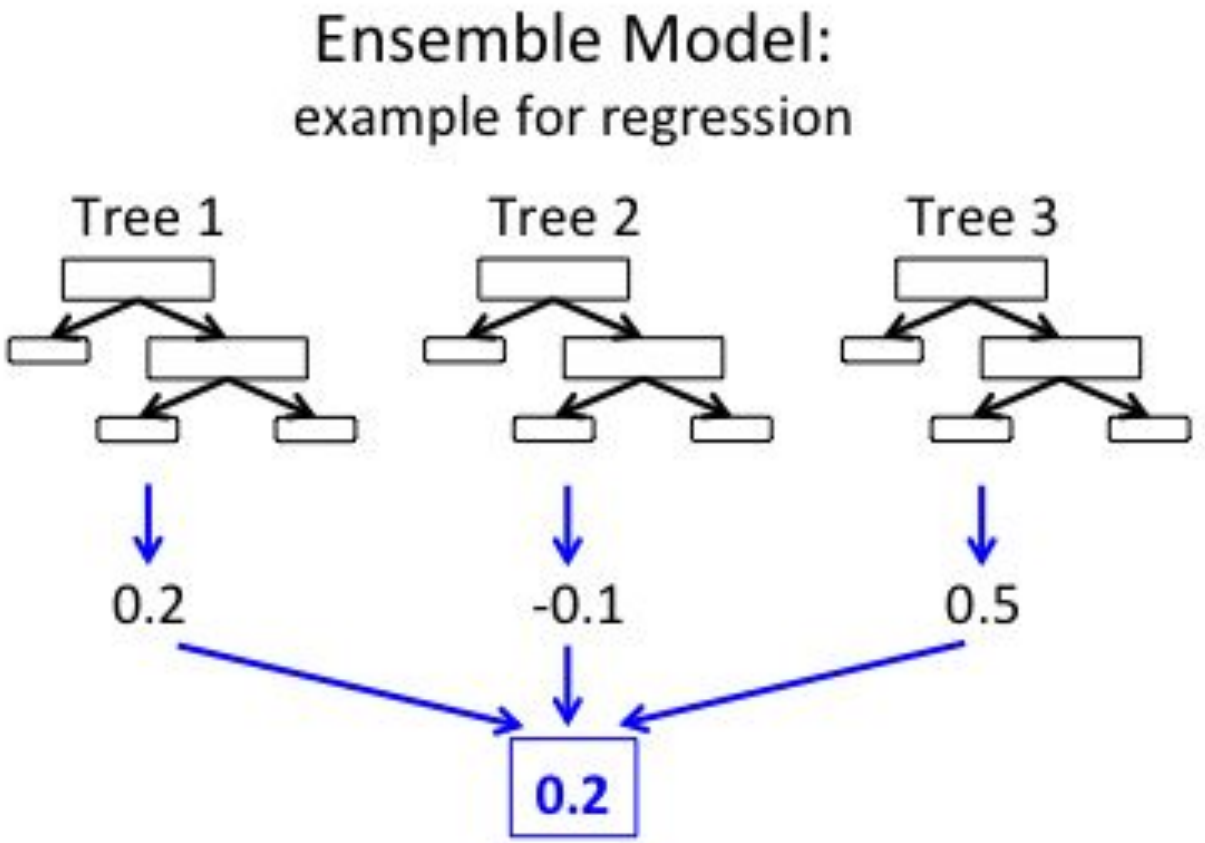


http://www.lewisgavin.co.uk/Machine-Learning-Decision-Tree/

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

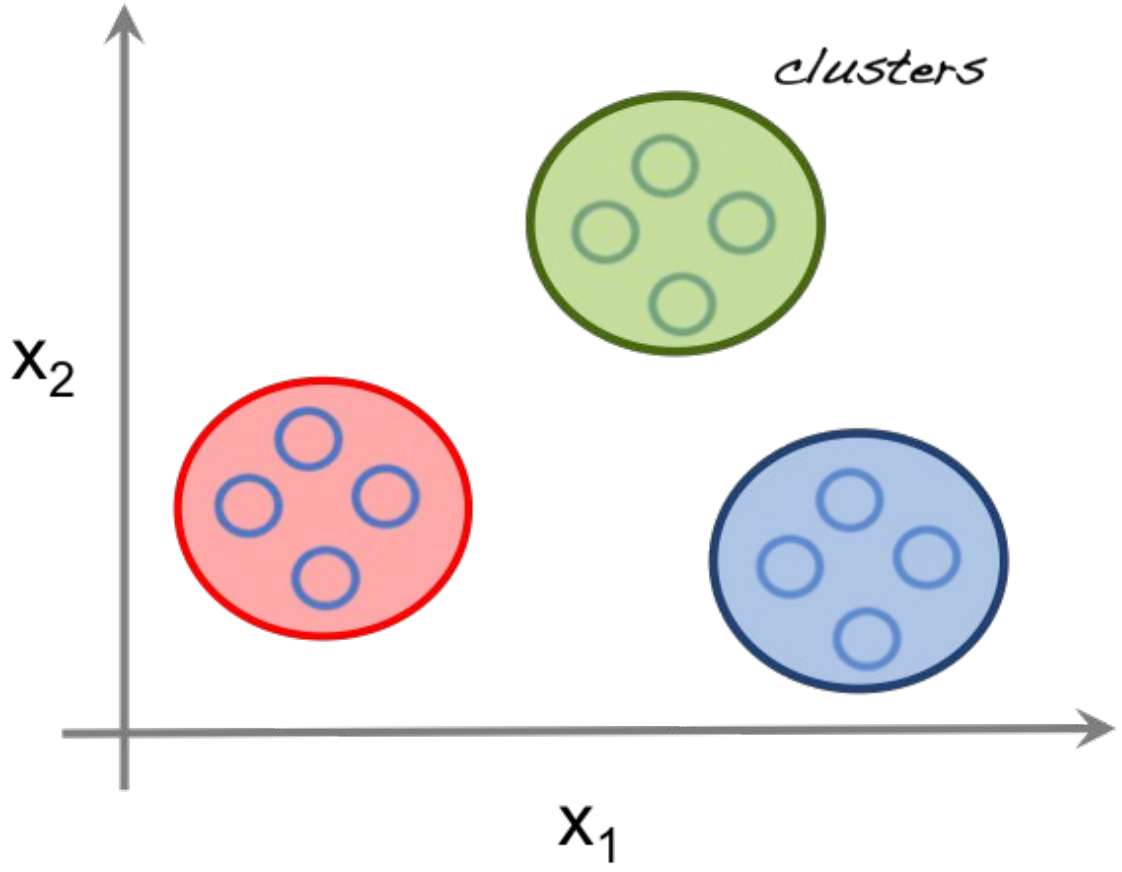
Notebook 3

Regression.ipynb

Unsupervised Learning

Unsupervised Learning

Search for data structures



Unsupervised Learning

data	features
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~~

There is NO ground truth!

Data generation model:

$x_i \sim P_\chi$

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

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

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

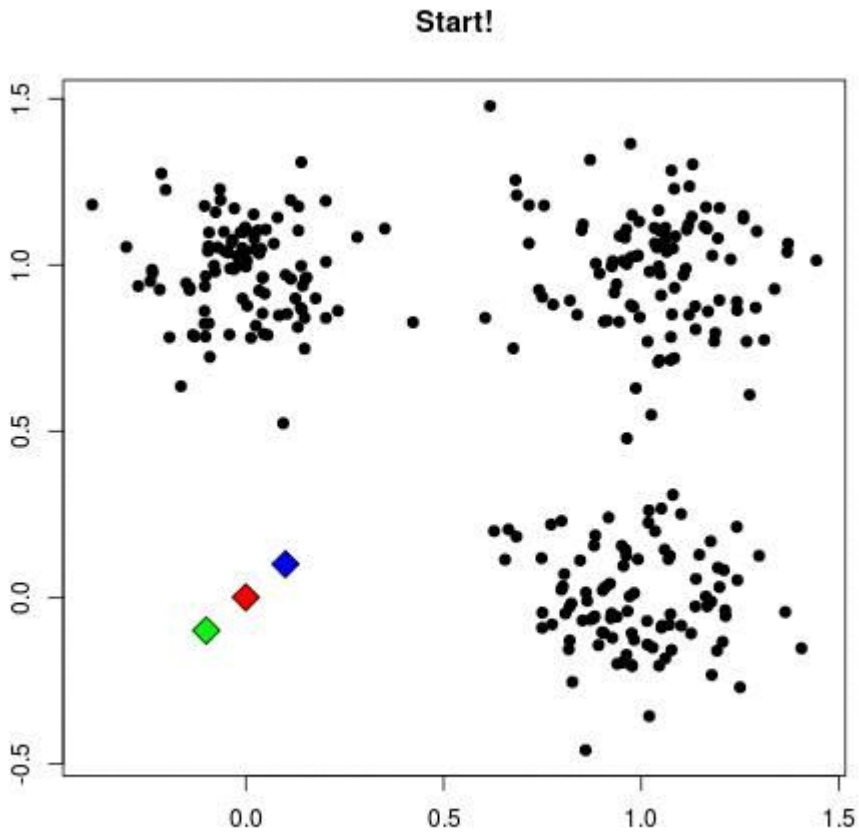
Clustering

1: assign centroids

2: each data point is attributed to the closest centroid

3: the centroid is moved to the center of the data points attributed to it

Repeat 2 and 3 until convergence

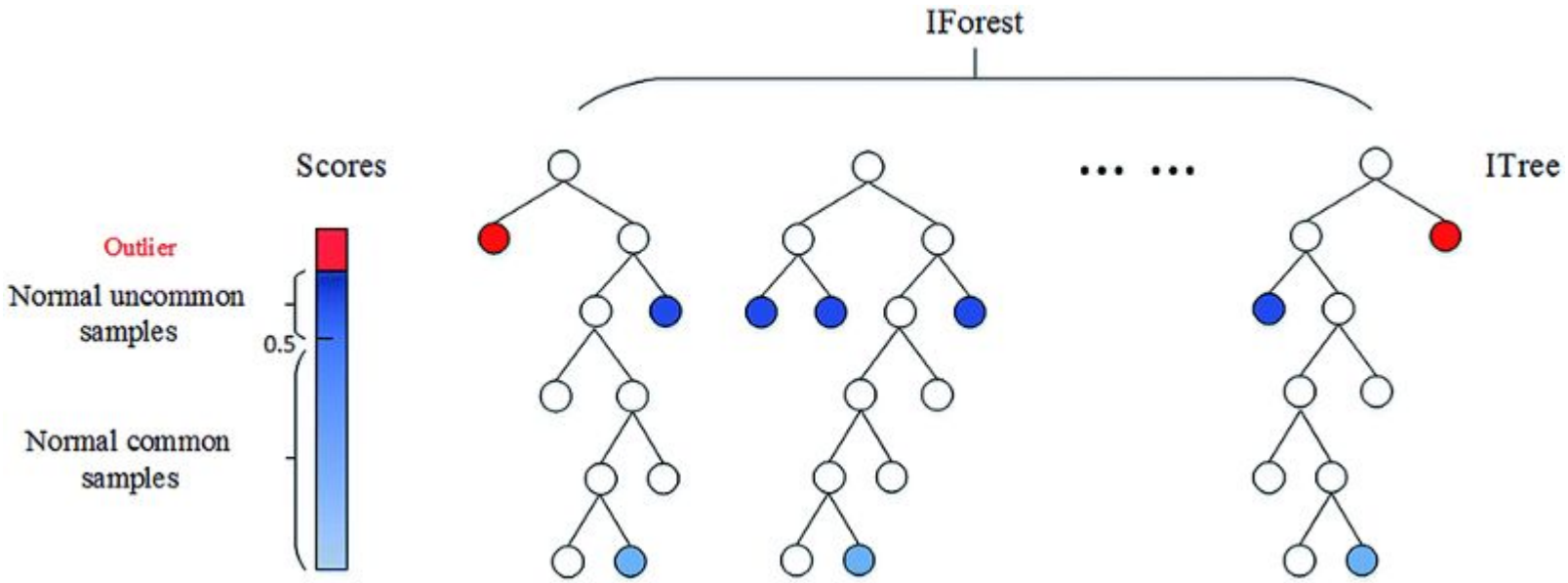


https://mubaris.com/2017/10/01/kmeans-clustering-in-python/

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

2. Supervised Machine Learning

$\Downarrow$

Minimize loss function

3. The Machine Learning algorithm, or learner, is only 1 element in a larger Machine Learning model

$\Longrightarrow$ you should pay attention to the other elements as well!

THANK YOU

Acknowledgement

- H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).