



The Bayesian optimization an active learning method for optimising experimental parameters

Sébastien Junier^{1,2}, Céline Barreteau¹ Christophe Bajan², Guillaume Lambard² and
Jean-Claude Crivello²

¹ *Univ Paris Est Créteil, CNRS, ICMPE, UMR 7182, 2 rue Henri Dunant, 94320 Thiais, France*

² *CNRS – Saint-Gobain – NIMS, Laboratory for Innovative Key Materials and Structures (LINK, IRL 3629), 1-1 Namiki, 305-0044 Tsukuba, Japan*

ICMMO seminar – Thursday the 7th of November, 2024



My background

Studies

- 1 Undergraduate's degree : Physics–Chemistry, Paris Saclay (2020)
- 2 Master's degree : Materials Science, Paris Saclay (2022)
- 3 PhD in ICMPE (3rd year) : Prediction of ESO by DFT and machine learning

Internships

- 1 CoAg nanoparticles : study by Monte Carlo simulation, ICMMO, Orsay, 4 months (2021)
- 2 Screening of ESO by DFT and machine learning, ICMPE, Thiais, 5 months (2022)
- 3 Accelerating ESO discovery using active learning, LINK, Tsukuba JAPAN, 3 months (2022)

Poster presentations in international congree

- International Conference on high-entropy Materials 2023
- STAC-13 and D2Mat symposium, Tsukuba, Japan, February 2024 (Outstanding Poster Award)

Outlook

- 1 Introduction
 - Machine Learning
 - Interest
- 2 Models
 - Regression model
 - Bayesian Optimisation
 - MADGUI: User Interface developed by NIMS
- 3 Theoretical Examples
 - One dimension
 - Two dimensions
- 4 Real examples
 - Application on SPS optimisation
- 5 Conclusion

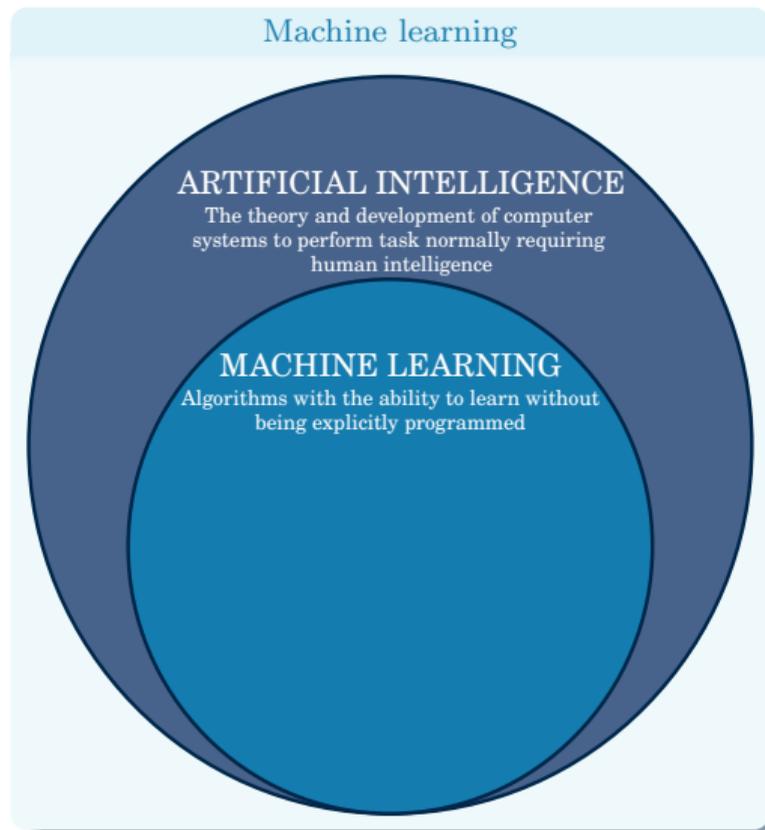
Introduction

Machine learning

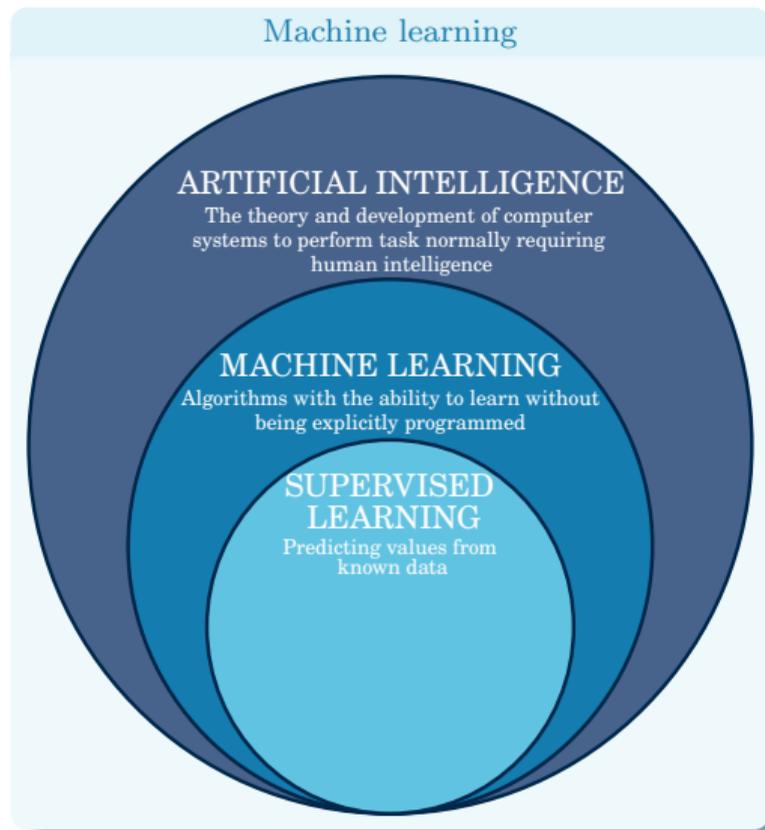
ARTIFICIAL INTELLIGENCE

The theory and development of computer systems to perform task normally requiring human intelligence

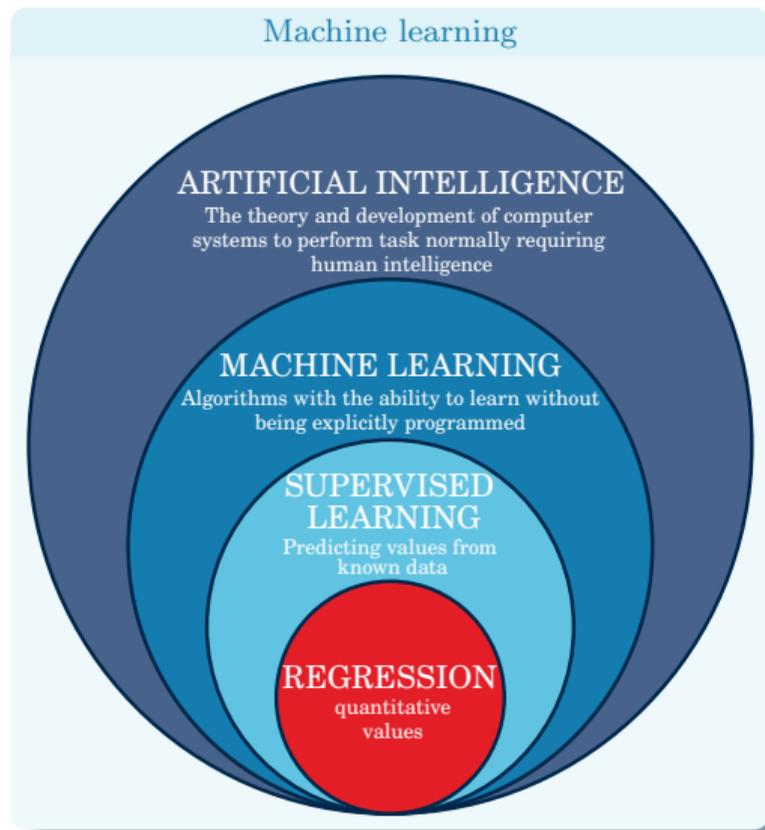
Introduction



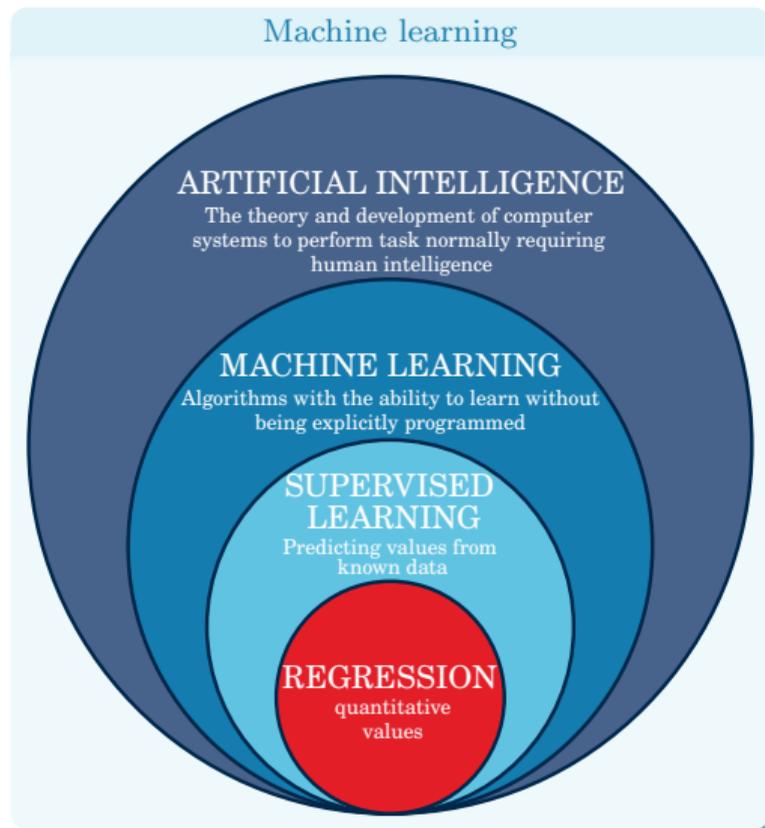
Introduction



Introduction



Introduction



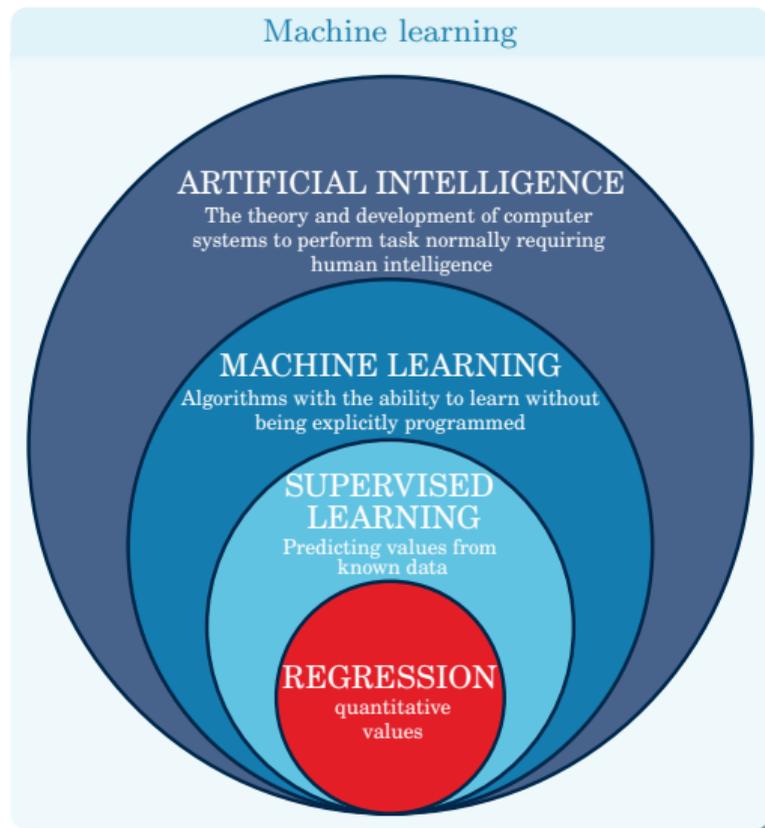
Regression

θ_1	θ_2	θ_3	y_1	y_2
θ_1^a	θ_2^a	θ_3^a	y_1^a	y_2^a
θ_1^b	θ_2^b	θ_3^b	y_1^b	y_2^b
θ_1^c	θ_2^c	θ_3^c	y_1^c	y_2^c
θ_1^d	θ_2^d	θ_3^d	y_1^d	y_2^d

$$\theta^i = \begin{pmatrix} \theta_1^i \\ \theta_2^i \\ \theta_3^i \end{pmatrix} \quad \mathbf{Y}^i = \begin{pmatrix} y_1^i \\ y_2^i \end{pmatrix}$$

From the table above, we look for the function
 $f : \theta \rightarrow \mathbf{Y}$

Introduction



Regression

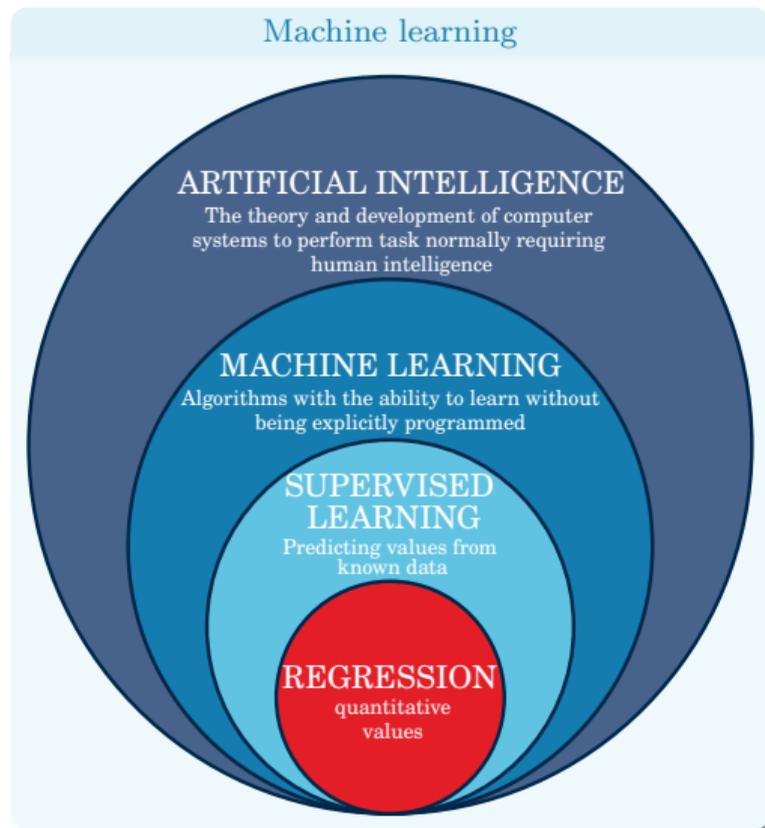
θ_1	θ_2	θ_3	y_1	y_2
θ_1^a	θ_2^a	θ_3^a	y_1^a	y_2^a
θ_1^b	θ_2^b	θ_3^b	y_1^b	y_2^b
θ_1^c	θ_2^c	θ_3^c	y_1^c	y_2^c
θ_1^d	θ_2^d	θ_3^d	y_1^d	y_2^d

$$\theta^i = \begin{pmatrix} \theta_1^i \\ \theta_2^i \\ \theta_3^i \end{pmatrix} \quad \mathbf{Y}^i = \begin{pmatrix} y_1^i \\ y_2^i \end{pmatrix}$$

From the table above, we look for the function
 $f: \theta \rightarrow \mathbf{Y}$

θ_1^e	θ_2^e	θ_3^e	?	?
θ_1^f	θ_2^f	θ_3^f	?	?

Introduction



Regression

θ_1	θ_2	θ_3	y_1	y_2
θ_1^a	θ_2^a	θ_3^a	y_1^a	y_2^a
θ_1^b	θ_2^b	θ_3^b	y_1^b	y_2^b
θ_1^c	θ_2^c	θ_3^c	y_1^c	y_2^c
θ_1^d	θ_2^d	θ_3^d	y_1^d	y_2^d

$$\theta^i = \begin{pmatrix} \theta_1^i \\ \theta_2^i \\ \theta_3^i \end{pmatrix} \quad \mathbf{Y}^i = \begin{pmatrix} y_1^i \\ y_2^i \end{pmatrix}$$

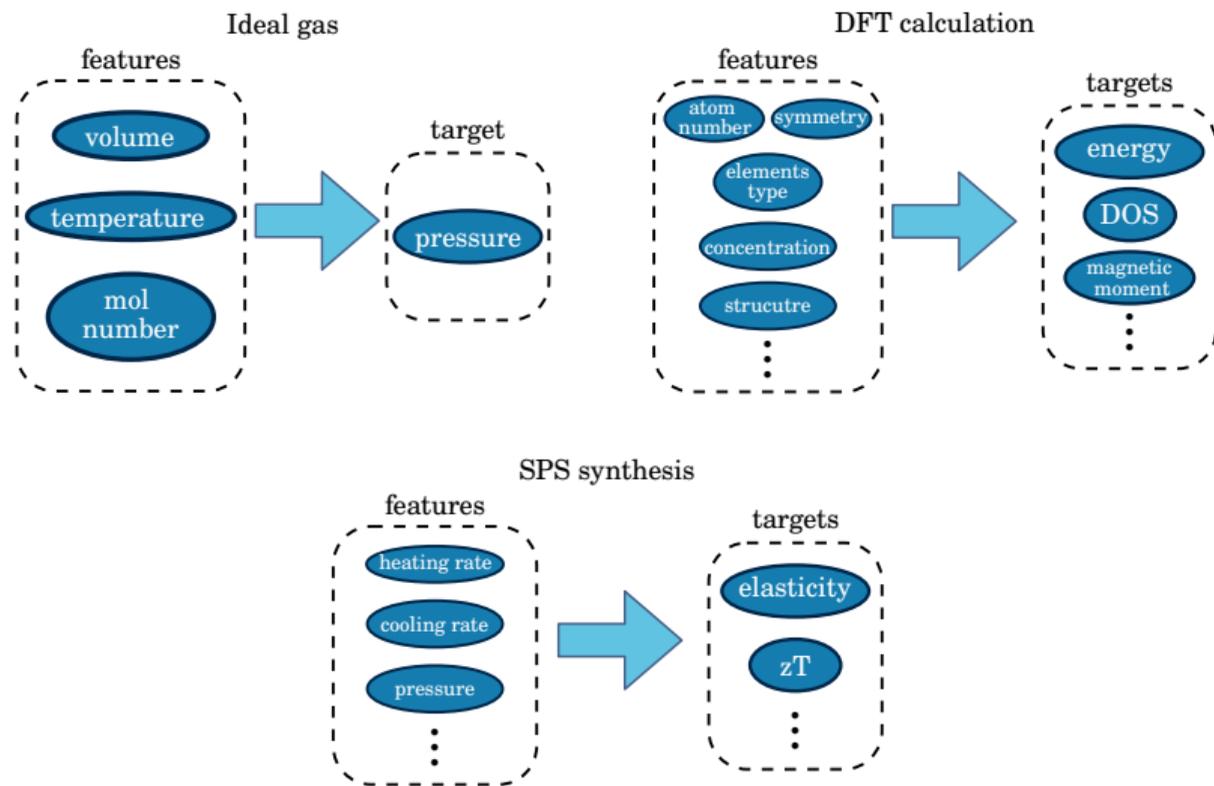
From the table above, we look for the function
 $f : \theta \rightarrow \mathbf{Y}$

θ_1^e	θ_2^e	θ_3^e	?	?
θ_1^f	θ_2^f	θ_3^f	?	?

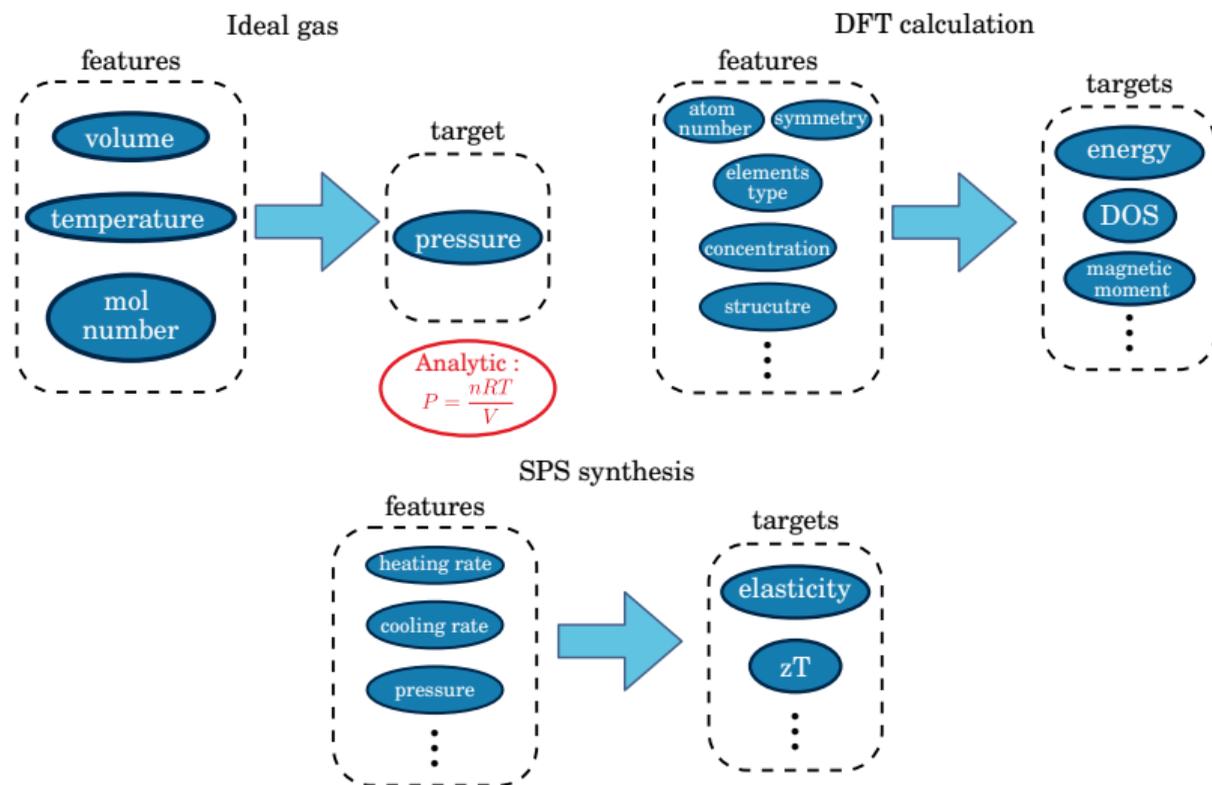
Formalism

θ : features \mathbf{Y} : targets

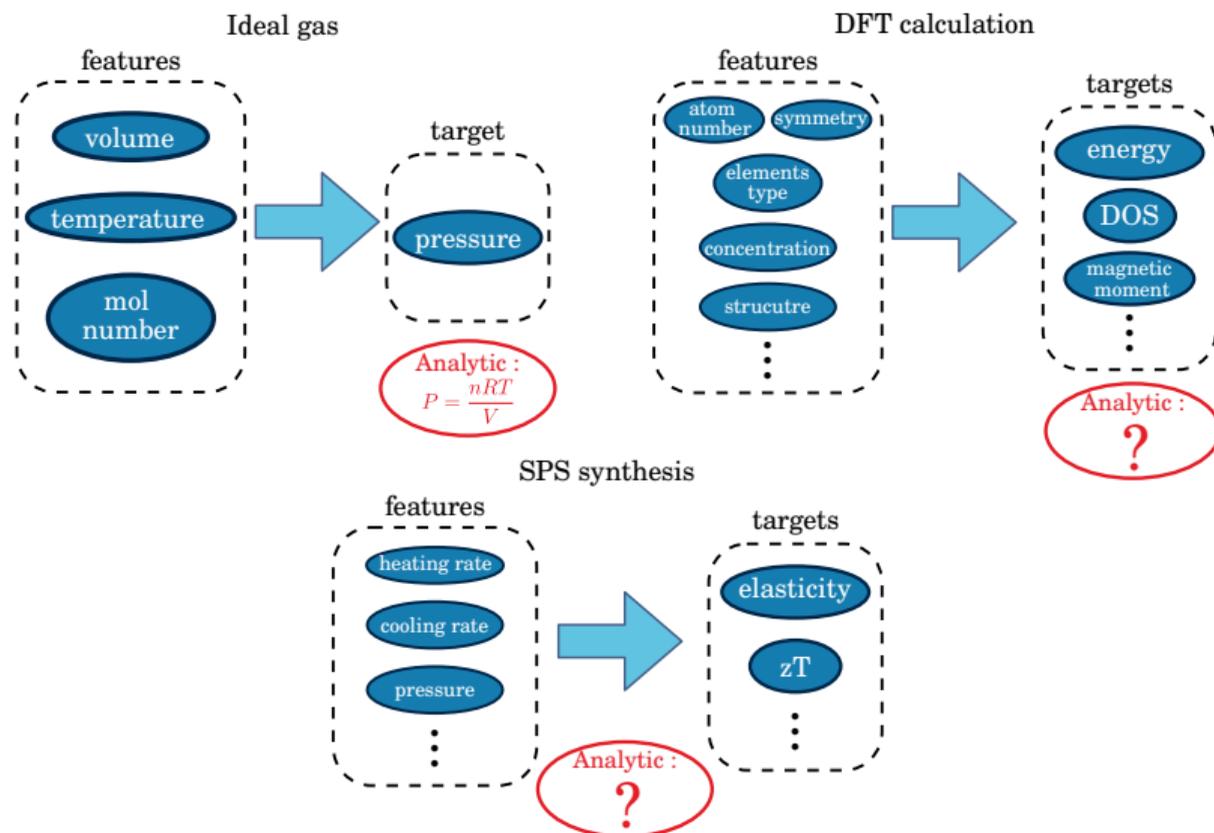
Some examples



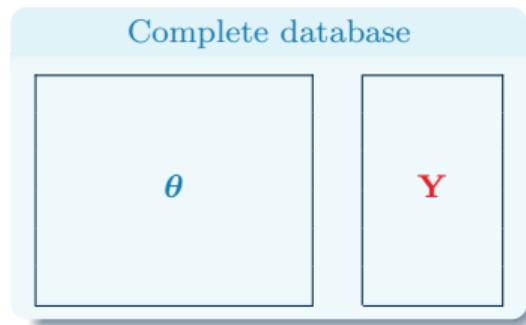
Some examples



Some examples



Regression Model (RM)



Regression Model (RM)

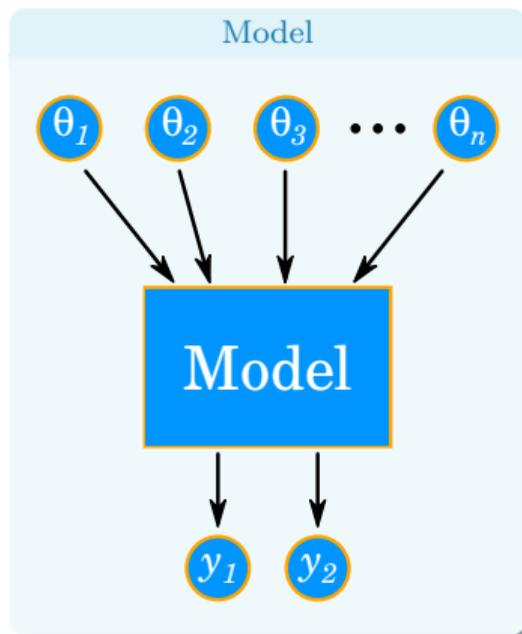
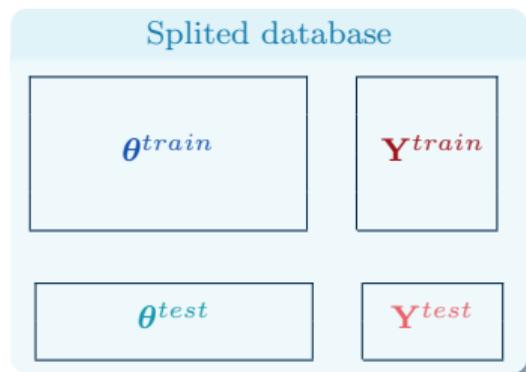
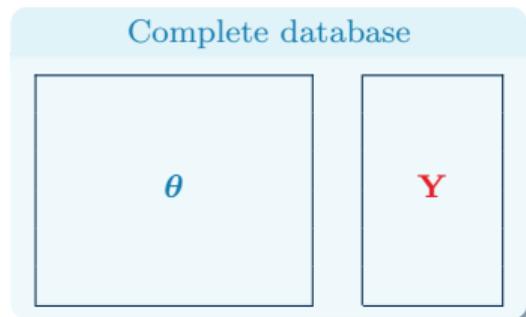
Complete database

 θ Y

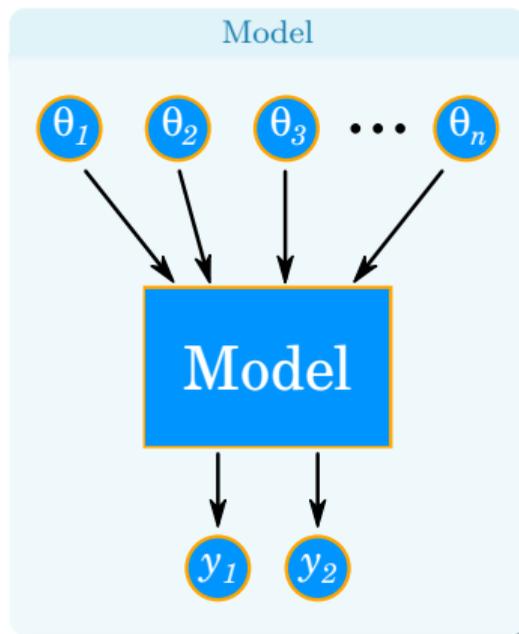
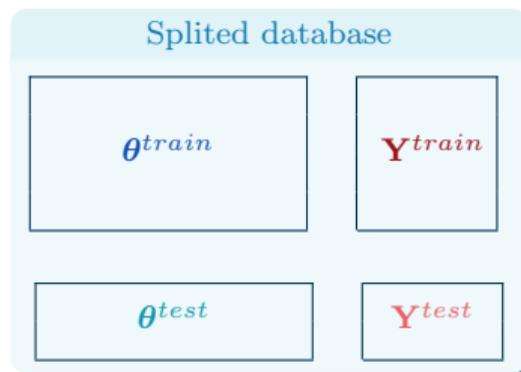
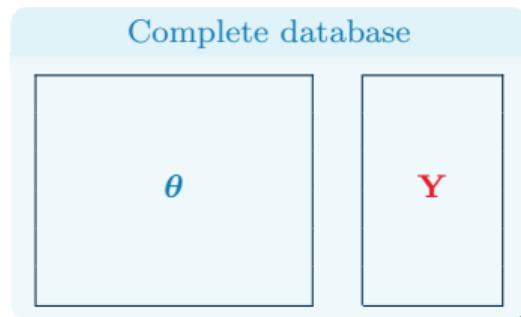
Splited database

 θ^{train} Y^{train} θ^{test} Y^{test}

Regression Model (RM)



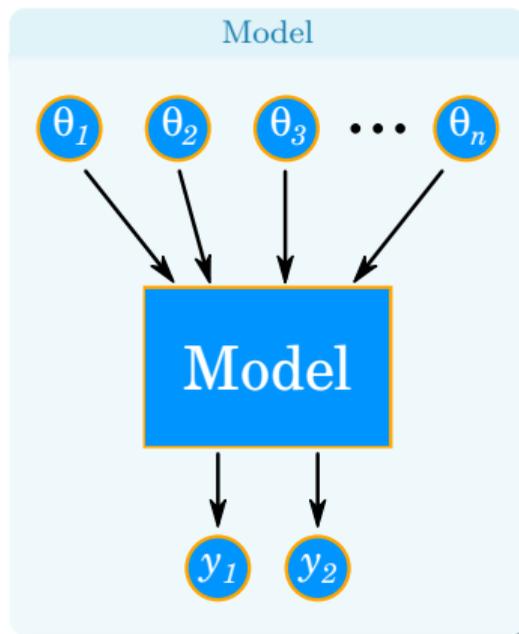
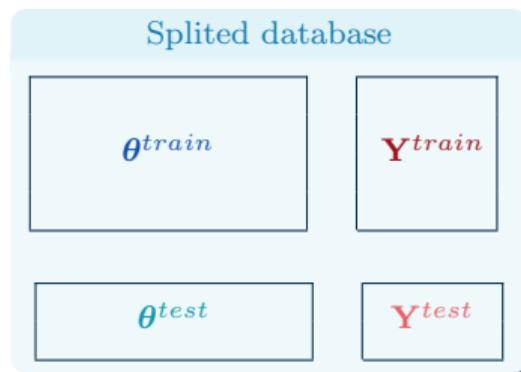
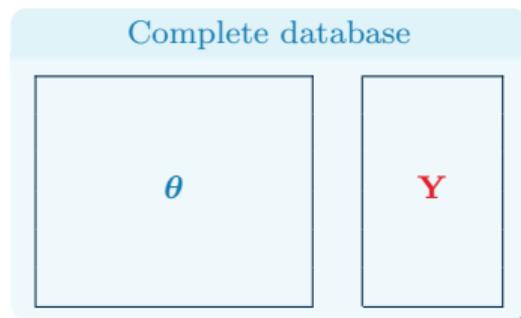
Regression Model (RM)



1 - Training

The model is trained on θ^{train} to fit Y^{train}

Regression Model (RM)



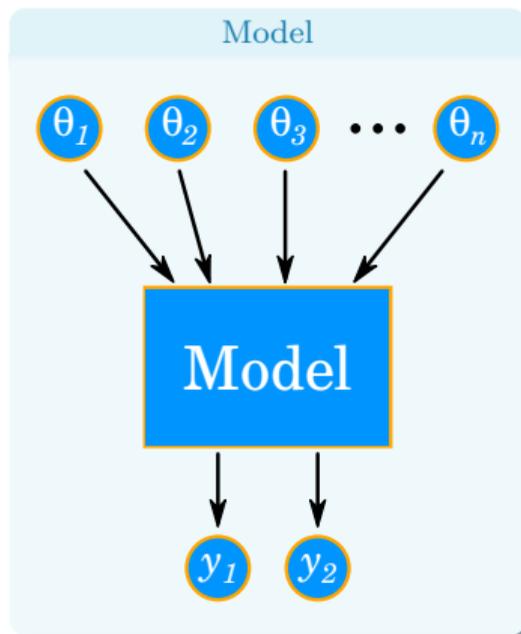
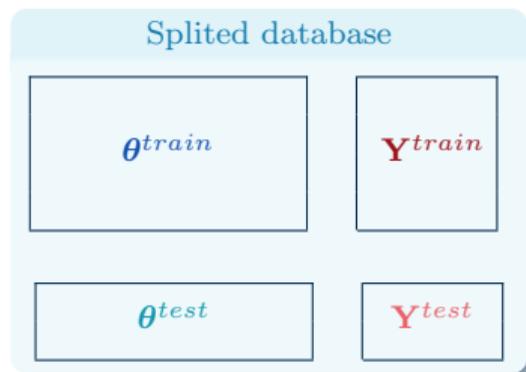
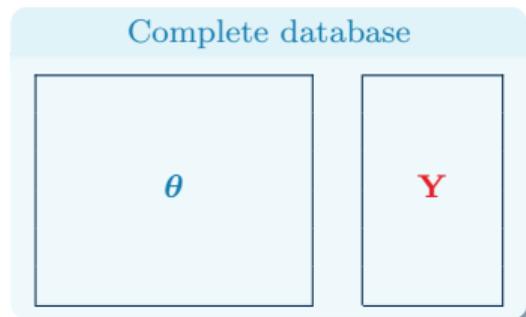
1 - Training

The model is trained on θ^{train} to fit Y^{train}

2 - Testing

The model is tested on θ^{test} and the results are compared with Y^{test}

Regression Model (RM)



1 - Training

The model is trained on θ^{train} to fit Y^{train}

2 - Testing

The model is tested on θ^{test} and the results are compared with Y^{test}

Database size

From 100 to several 1000 points

Bayesian Optimisation (BO)

Aim

Find features θ^i which correspond to an optimum of targets \mathbf{Y}

M. Krasser, "Bayesian Optimization", github (2018)

Bayesian Optimisation (BO)

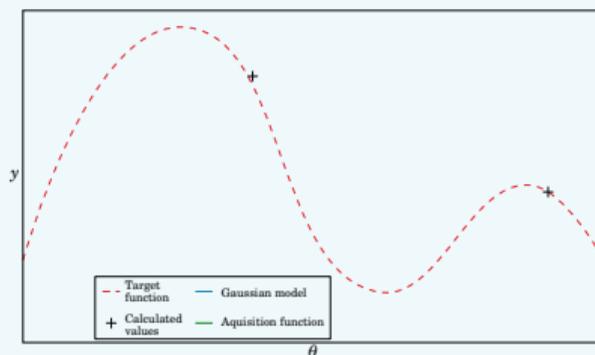
Aim

Find features θ^i which correspond to an optimum of targets \mathbf{Y}

Gaussian Process

Database
 $\{\theta\}=\{\mathbf{Y}\}$

Graphic Gaussian Process



Bayesian Optimisation (BO)

Aim

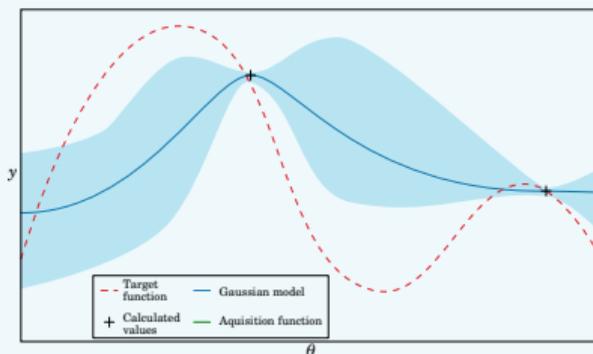
Find features θ^i which correspond to an optimum of targets \mathbf{Y}

Gaussian Process

Database
 $\{\theta\}=\{\mathbf{Y}\}$



Graphic Gaussian Process



Bayesian Optimisation (BO)

Aim

Find features θ^i which correspond to an optimum of targets \mathbf{Y}

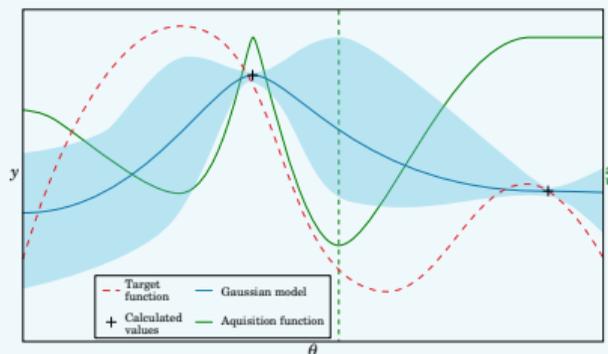
Gaussian Process

Database
 $\{\theta\}=\{\mathbf{Y}\}$



suggest
new θ

Graphic Gaussian Process

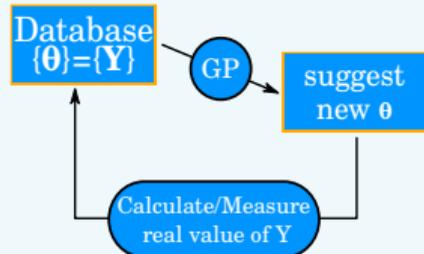


Bayesian Optimisation (BO)

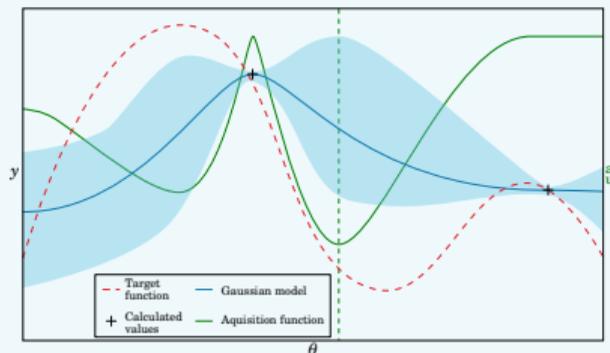
Aim

Find features θ^i which correspond to an optimum of targets \mathbf{Y}

Gaussian Process



Graphic Gaussian Process

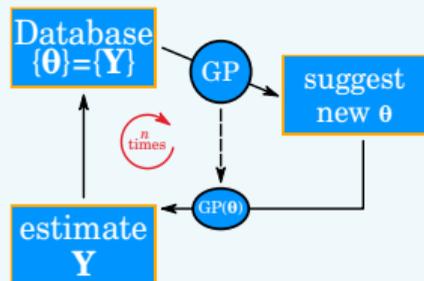


Bayesian Optimisation (BO)

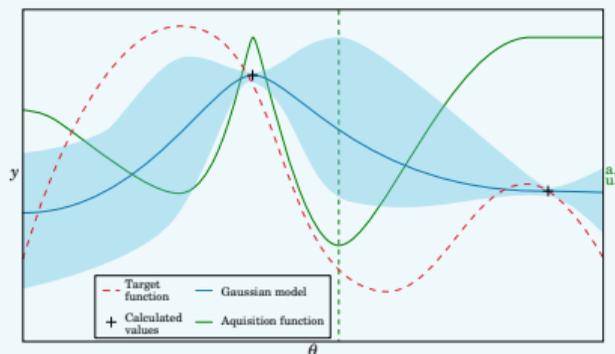
Aim

Find features θ^i which correspond to an optimum of targets \mathbf{Y}

Gaussian Process



Graphic Gaussian Process

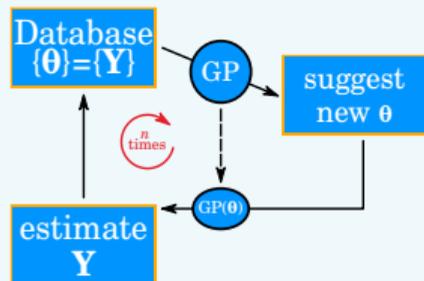


Bayesian Optimisation (BO)

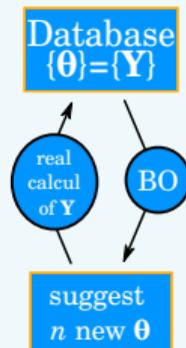
Aim

Find features θ^i which correspond to an optimum of targets \mathbf{Y}

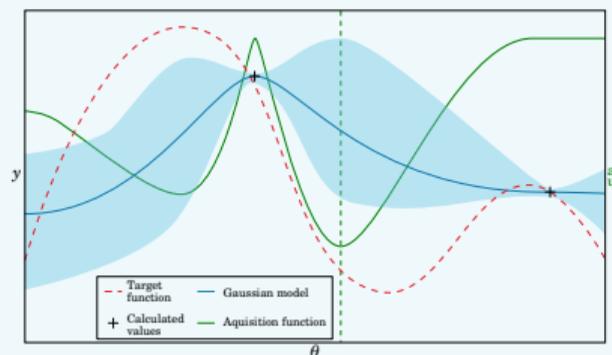
Gaussian Process



BO framework



Graphic Gaussian Process

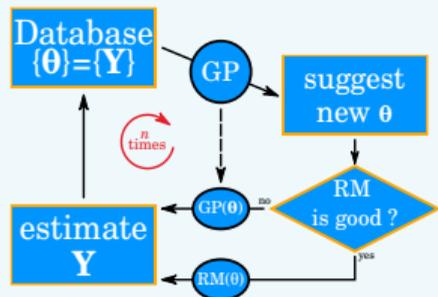


Bayesian Optimisation (BO)

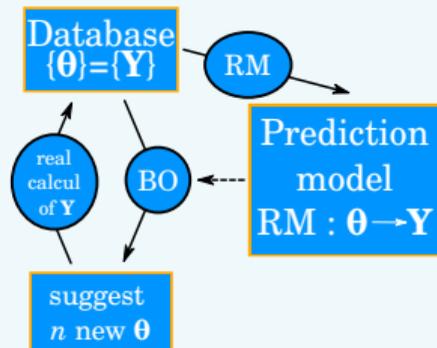
Aim

Find features θ^i which correspond to an optimum of targets \mathbf{Y}

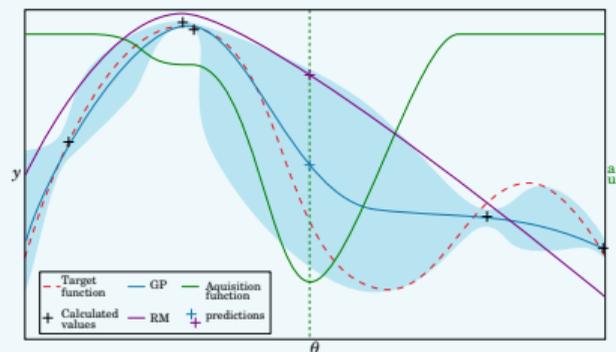
Gaussian Process



BO framework



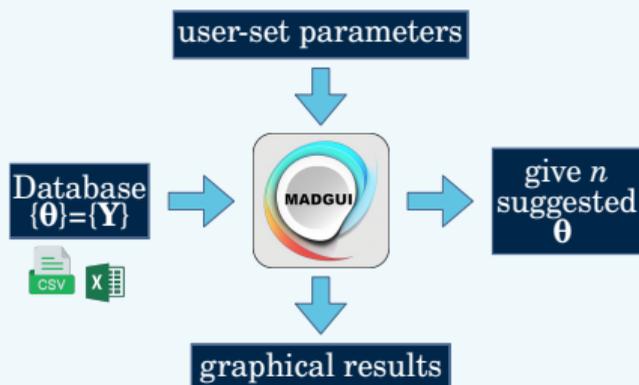
Graphic Gaussian Process



MADGUI

Fully graphical interface developed by [C. Bajan](#) and [G. Lambard](#) to makes BO accessible to researchers without extensive programming experience.

Global use



- Github : github.com/Lambard-ML-Team/MADGUI
- WEB app : lambard-ml-team-madgui.streamlit.it
- pre-print article : dx.doi.org/10.2139/ssrn.4855240

WEB application

MADGUI - Multi-Application Design GUI

Welcome! The objective of this project is to help you for data analysis, prediction model and to find the best next sample to reach your objective using Bayesian Optimisation. You will be asked to complete different parts of this program. You must start by completing the Main Page where you are currently located. Then you will be able to do either the prediction or the Bayesian optimization. Take note that if you change anything after submitting your selection, you must click "Submit" again.

1 - Data upload

Read carefully before uploading your data

△ Read before continuing △

Firstly, you will have to upload your data. To do that you have to click on the "Browse File" button in the sidebar.

Your data from the file data-zf.csv are represented below, you can check if everything is alright.

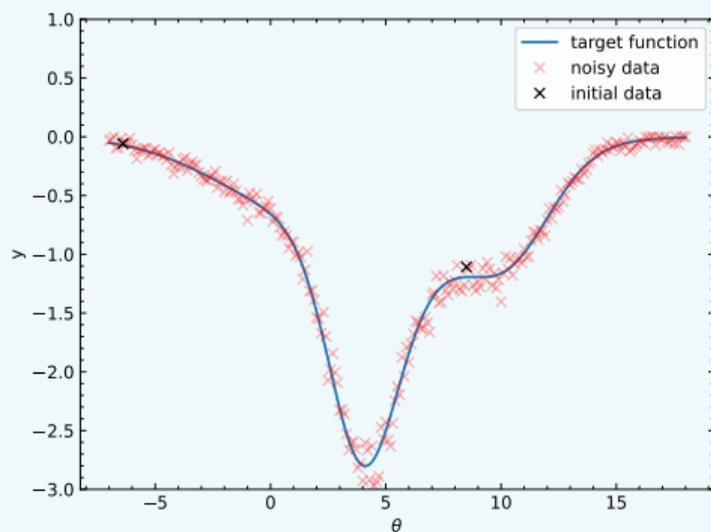
	Sample	HR	Sint	Stepit	CR	UP	mm	zt
0	1	300	773	15	300	50	14.76	0.276
1	2	300	973	15	300	50	15.31	0.269
2	3	50	773	15	50	50	15.24	0.27
3	4	300	773	15	300	50	15.32	0.269
4	5	175	773	15	170	50	15.14	0.271

Optimisation of one function

Global objective

Find the minimum y

Graphical representation

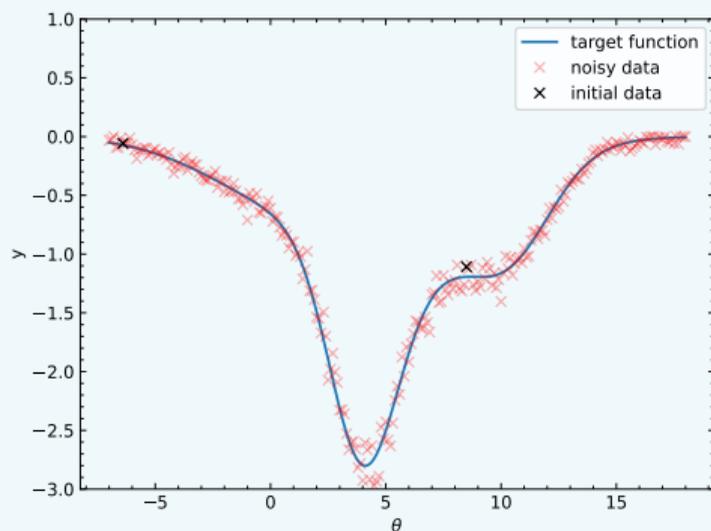


Optimisation of one function

Step

Import file : numeric table

Graphical representation



How it works on MADGUI



Navigation

Main Page

Prediction

Bayesian

About

Contact

Upload your data

Drag and drop file here

Limit 200MB per file • CSV, XLSX

Browse files

gausn0.csv
75.0B

MADGUI - Multi-Application Design GUI

Welcome! The objective of this project is to help you for data analysis, prediction model and to find the best next sample to reach your objective using Bayesian Optimisation. You will be asked to complete different parts of this program. You must start by completing the Main Page where you are currently located. Then you will be able to do either the prediction or the Bayesian optimization. Take note that if you change anything after submitting your selection, you must click 'Submit' again.

1 - Data upload

Read carefully before uploading your data

△ Read before continuing △

Firstly, you will have to upload your data. To do that you have to click on the "Browse File" button in the sidebar.

Your data from the file gausn0.csv are represented below, you can check if everything is alright.

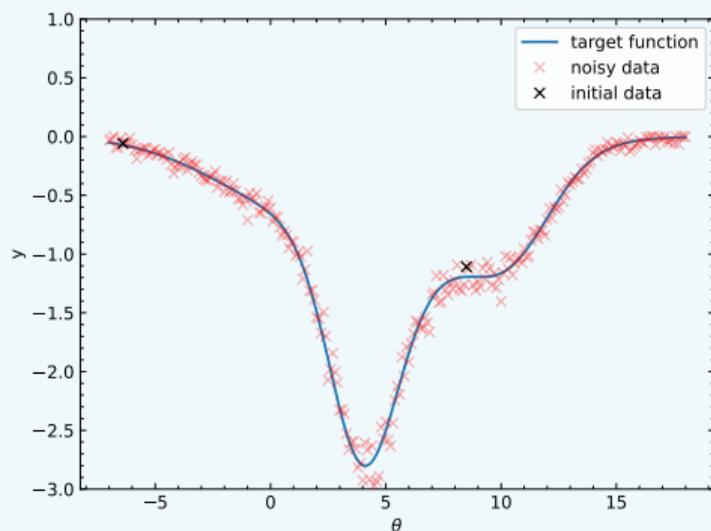
	point number	theta	y
0	1	8.5	-1.1083
1	2	-6.4	-0.0575

Optimisation of one function

Step

Select features and targets

Graphical representation



How it works on MADGUI



Navigation

Main Page

Prediction

Bayesian

About

Contact

Upload your data

Drag and drop file here

Limit 200MB per file • CSV, XLSX

Browse files

gausz0.csv
75.0B

alright.

	point number	theta	y
0	1	8.5	-1.1083
1	2	-6.4	-0.0575

2 - Selection of your features and targets for the project

On this section of the program, you must select which columns of your dataset are the features you want to analyse and which columns are the targets that you want to predict or improve.

Columns with a standard deviation of 0 are already deselected, as well as columns that contain text. This is done automatically to eliminate columns that do not provide useful information for analysis and prediction.

Features - Unselected the one you don't need:

point number × theta × y ×

Select your targets to: submit button:

Choose an option

Submit

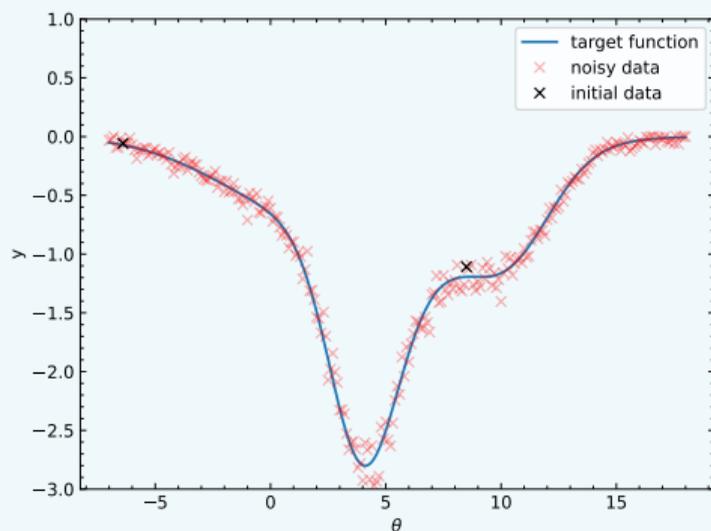
Submit your selection.

Optimisation of one function

Step

We have an analysis of data

Graphical representation



How it works on MADGUI

The screenshot shows the MADGUI web interface. On the left is a navigation menu with options: Main Page (highlighted), Prediction, Bayesian, About, and Contact. Below the menu is a file upload section for 'gausz0.csv' (75.0B). On the right, there are input fields for 'theta' and 'y', a 'Submit' button, and a table showing the selected data:

theta	y
0	8.5 -1.1083
1	-6.4 -0.0575

Below the table is a '3 - Quick analysis' section with a summary table:

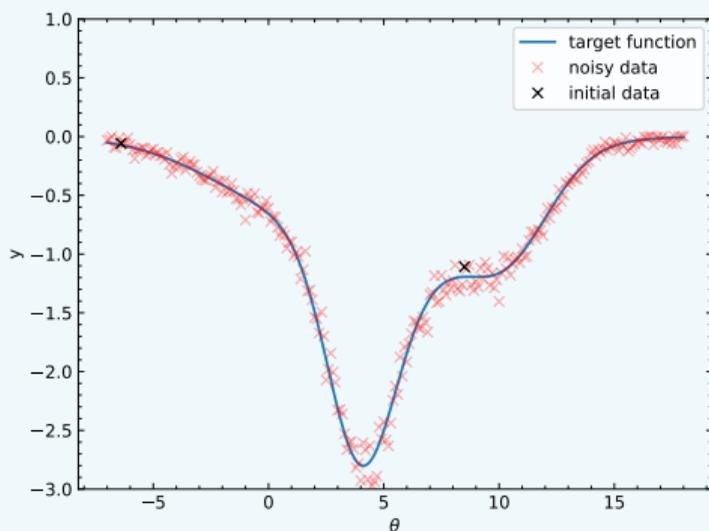
	theta	y
count	2	2
mean	1.05	-0.5829
std	10.5359	0.743
min	-6.4	-1.1083
25%	-2.675	-0.8456
50%	1.05	-0.5829
75%	4.775	-0.3202
max	8.5	-0.0575

Optimisation of one function

Step

Choose domain of feature

Graphical representation



How it works on MADGUI



	Min	Max	Step	Specific Values
theta	0	0	0	0

Click below to download a CSV. Fill out the template with the boundaries and step of the features you selected before. Then you will have to upload it below

[Click here to download the file with your features](#)

Bayesian Optimisation

Limits and constraints

	Min	Max	Step	Specific Values
theta	-7	18	0.1	0

[Validate limits](#)

Upload the file containing the limits that you fixed

Select the file that you have completed

Drag and drop file here
Limit: 200MB per file • CSV

[Browse files](#)

Save your limits for next time

Make sure that everything is correct

With the limits that you choose, the possible values for each feature are:

	name	type	domain
0	theta	discrete	-7 -6.9 -6.8000000000000001 -6.7000000000000001 -6.6000000000000001

The number of possibilities are : 2.51e+02

Upload your data

Drag and drop file here
Limit: 200MB per file • CSV, XLSX

[Browse files](#)

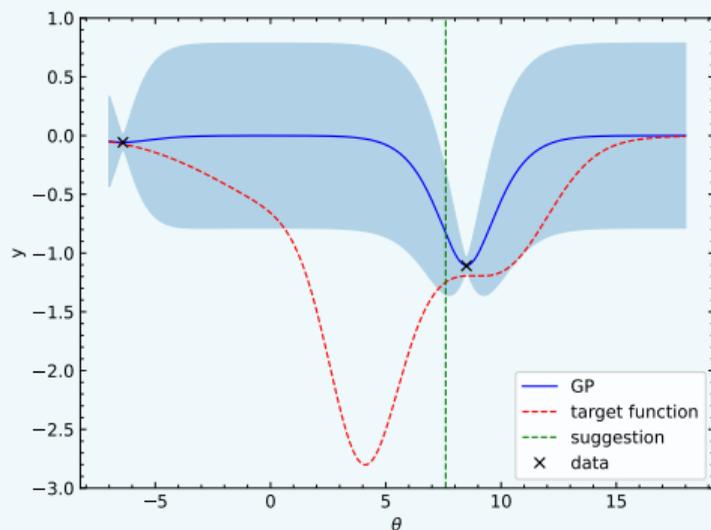
gausn0.csv
75.0B

Optimisation of one function

Step

Select parameters of BO

Graphical representation



How it works on MADGUI

The screenshot shows the MADGUI web interface. The main heading is "Optimisation". Below it, there is a description of the Acquisition function. The interface includes a navigation menu with options: Main Page, Prediction, Bayesian (highlighted), About, and Contact. The main form has the following fields:

- How many target do you want to optimise? (max 3): 1
- Select which target you want to optimise, be careful to not choose the same target twice: y
- Select if you want to maximize/minimize this target: minimize
- Select the number of sample that you want: 1
- Execute the Bayesian Optimisation (button)
- The result of the bayesian optimisation are:

theta
0
7.7
- Click below to download the CSV with the proposed data for optimisation. (button)

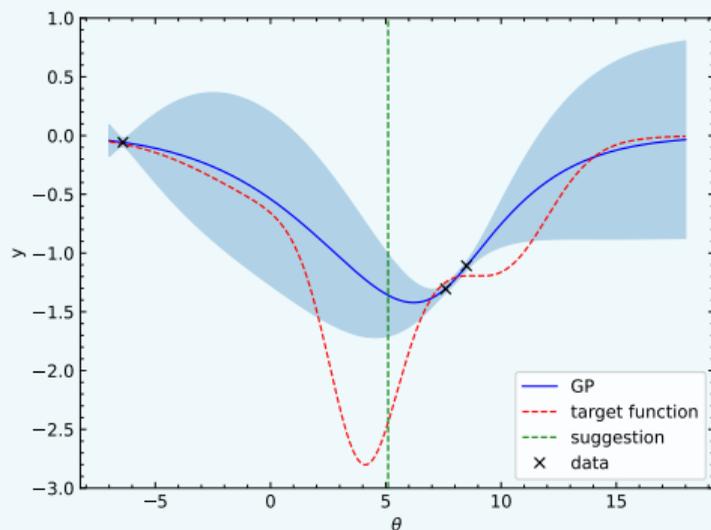
At the bottom, there is a file upload section for "gausn0.csv" (75.0B) and a "Click here" button.

Optimisation of one function

Step

Calculate y , add to in database and repeat

Graphical representation



How it works on MADGUI



point number	theta	y
0	1	8.5 -1.1083
1	2	-6.4 -0.0575
2	3	7.6 -1.3051

2 - Selection of your features and targets for the project

On this section of the program, you must select which columns of your dataset are the features you want to analyse and which columns are the targets that you want to predict or improve.

Columns with a standard deviation of 0 are already deselected, as well as columns that contain text. This is done automatically to eliminate columns that do not provide useful information for analysis and prediction.

Features - Unselected the one you don't need :

theta x

Select your targets then click the submit button :

y x

Submit

The data that you selected (Features + Targets) are :

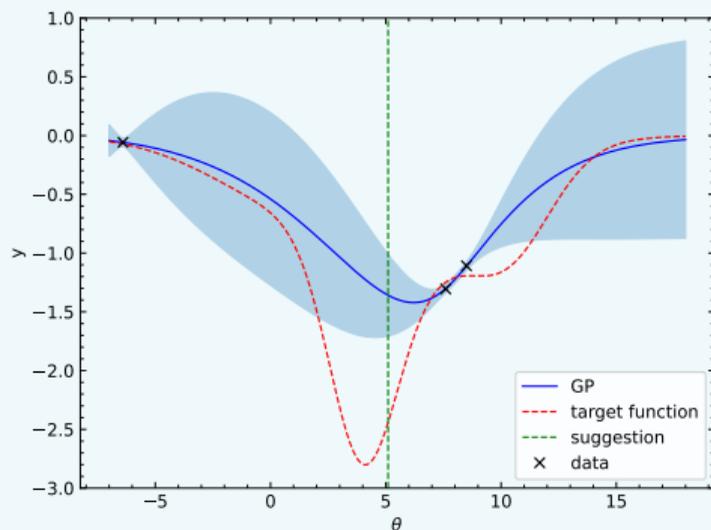
	theta	y
0	8.5	-1.1083
1	-6.4	-0.0575

Optimisation of one function

Step

Calculate y , add to in database and repeat

Graphical representation



How it works on MADGUI

The screenshot shows the MADGUI web interface. The main content area is titled "Optimisation" and contains the following settings:

- Optimisation:** The Acquisition function is Expected Improvement.
- How many target do you want to optimise? (max 3):** 1
- Select which target you want to optimise, be careful to not choose the same target twice:** y
- Select if you want to maximize/minimize this target:** minimize
- Select the number of sample that you want:** 1
- Execute the Bayesian Optimisation:** [Button]
- The result of the bayesian optimisation are:**
 - theta: 0 5.1
- Click below to download the CSV with the proposed data for optimisation:** [Click here]

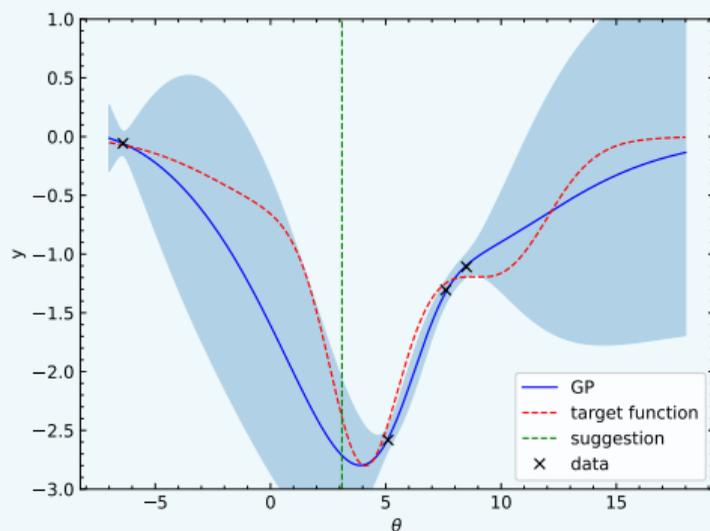
The interface also features a navigation menu with options: Main Page, Prediction, Bayesian (highlighted), About, and Contact. A file upload section is visible at the bottom, showing a file named "gausn1.csv" (101.0B) with a "Browse files" button.

Optimisation of one function

Step

Calculate y , add to in database and repeat

Graphical representation



How it works on MADGUI



Optimisation

The Acquisition function of this bayesian optimization is Expected Improvement, it is used to balance exploration and exploitation by quantifying the expected improvement of a candidate point over the current best observed value in the optimization process. It helps guide the search for the global optimum in an efficient manner.

How many target do you want to optimise? (max 3)

1

Select which target you want to optimise, be careful to not choose the same target twice

y

Select if you want to maximize/minimize this target

minimize

Select the number of sample that you want

1

1

20

Execute the Bayesian Optimisation

The result of the bayesian optimisation are :

theta

0 5.1

Click below to download the CSV with the proposed data for optimisation.

Click here

Upload your data

Drag and drop file here

Limit 200MB per file • CSV, XLSX

Browse files

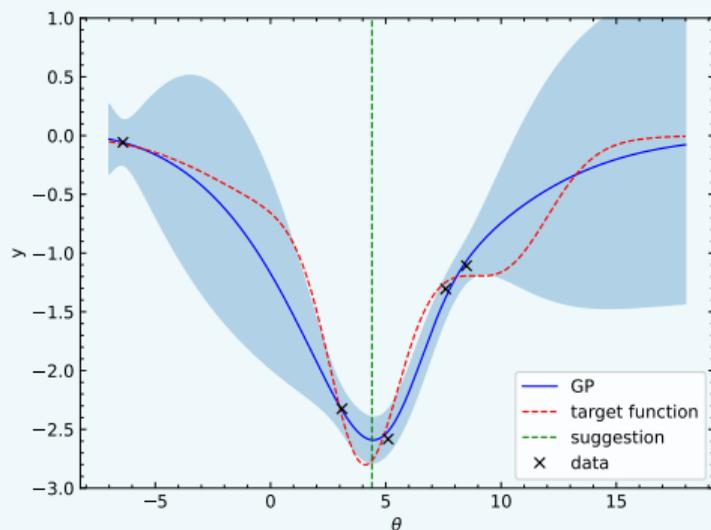
gausn1.csv
101.0B

Optimisation of one function

Step

Calculate y , add to in database and repeat

Graphical representation



How it works on MADGUI

The screenshot shows the MADGUI web interface. At the top left is the MADGUI logo. Below it is a navigation menu with options: Main Page, Prediction, Bayesian (highlighted in green), About, and Contact. The main content area is titled 'Optimisation' and contains the following information:

- Optimisation**: The Acquisition function of this bayesian optimization is Expected Improvement, it is used to balance exploration and exploitation by quantifying the expected improvement of a candidate point over the current best observed value in the optimization process. It helps guide the search for the global optimum in an efficient manner.
- How many target do you want to optimise? (max 3)**: A dropdown menu set to '1'.
- Select which target you want to optimise, be careful to not choose the same target twice**: A dropdown menu set to 'y'.
- Select if you want to maximize/minimize this target**: A dropdown menu set to 'minimize'.
- Select the number of sample that you want**: A slider set to '1'.
- Execute the Bayesian Optimisation**: A button.
- The result of the bayesian optimisation are :** A table showing the result:

theta
0 5.1
- Click below to download the CSV with the proposed data for optimisation.**: A 'Click here' button.

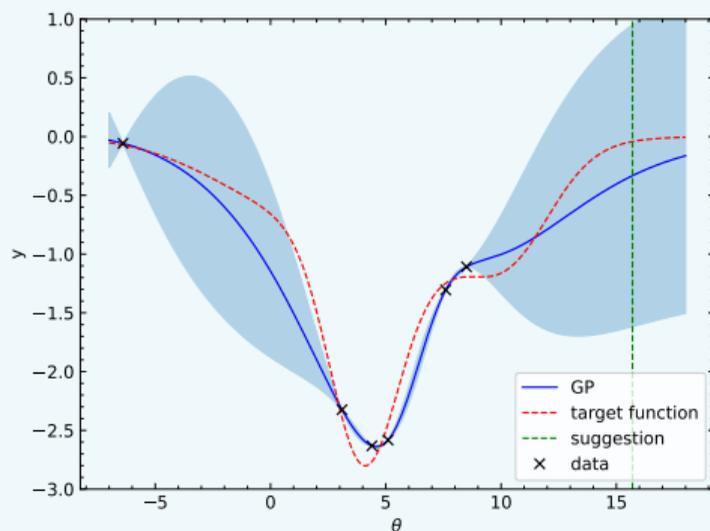
At the bottom left, there is a section for 'Upload your data' with a file upload area and a 'Browse files' button. A file named 'gausn1.csv' (101.0B) is shown as uploaded.

Optimisation of one function

Step

Calculate y , add to in database and repeat

Graphical representation



How it works on MADGUI

The screenshot shows the MADGUI web interface. The main content area is titled "Optimisation" and contains the following configuration options:

- Optimisation:** The Acquisition function is Expected Improvement, used to balance exploration and exploitation.
- How many target do you want to optimise? (max 3):** 1
- Select which target you want to optimise, be careful to not choose the same target twice:** y
- Select if you want to maximize/minimize this target:** minimize
- Select the number of sample that you want:** 1
- Execute the Bayesian Optimisation:** A button to start the optimization process.
- The result of the bayesian optimisation are:** A table showing the result for the parameter theta, with a value of 0.51.
- Click below to download the CSV with the proposed data for optimisation:** A button labeled "Click here".

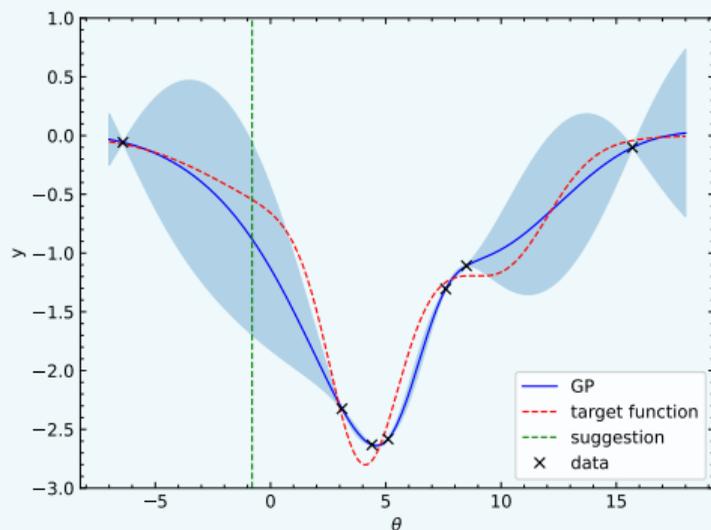
The interface also includes a navigation menu with options: Main Page, Prediction, Bayesian (highlighted), About, and Contact. There is also a section for uploading data, with a file named "gausn1.csv" (101.0B) already uploaded.

Optimisation of one function

Step

Calculate y , add to in database and repeat

Graphical representation



How it works on MADGUI

The screenshot shows the MADGUI web interface. The main content area is titled "Optimisation" and contains the following elements:

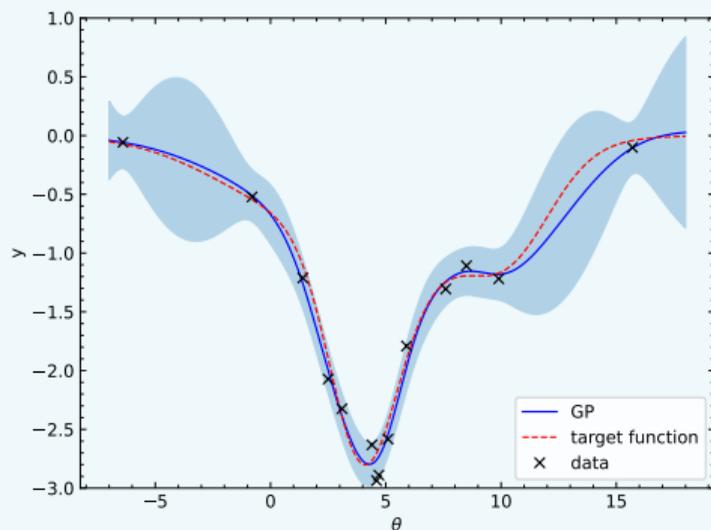
- Navigation:** A sidebar menu with options: Main Page, Prediction, Bayesian (highlighted), About, and Contact.
- Optimisation Form:**
 - How many target do you want to optimise? (max 3):
 - Select which target you want to optimise, be careful to not choose the same target twice:
 - Select if you want to maximize/minimize this target:
 - Select the number of sample that you want:
 -
- Data Upload:**
 - Upload your data: Drag and drop file here (Limit 200MB per file • CSV, XLSX)
 -
 - File: gauzn1.csv (101.0B)
- Results:**
 - The result of the bayesian optimisation are:
 - Parameter: theta
 - Value: 0 5.1
 - to download the CSV with the proposed data for optimisation.

Optimisation of one function

Step

Calculate y , add to in database and repeat

Graphical representation



How it works on MADGUI

The screenshot shows the MADGUI web interface. The navigation menu on the left includes: Main Page, Prediction, Bayesian (highlighted), About, and Contact. The main content area is titled "Optimisation" and contains the following settings:

- How many target do you want to optimise? (max 3): 1
- Select which target you want to optimise, be careful to not choose the same target twice: y
- Select if you want to maximize/minimize this target: minimize
- Select the number of sample that you want: 1

There is an "Execute the Bayesian Optimisation" button. Below this, the result of the optimization is shown as a table:

theta
0
5.1

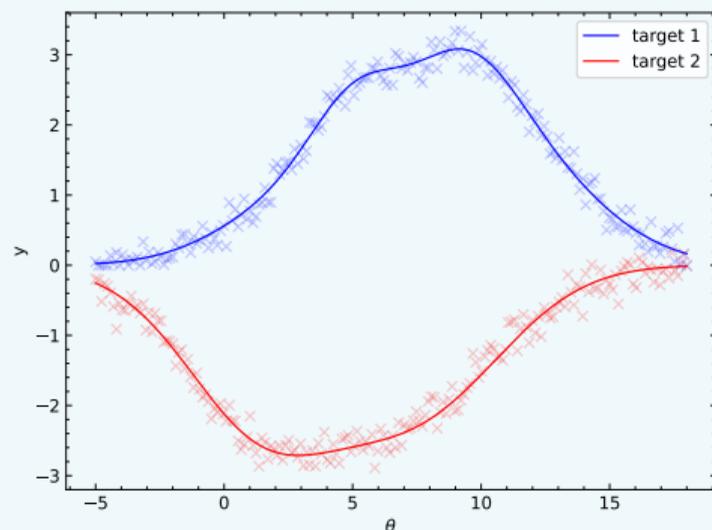
At the bottom, there is a section for uploading data, with a file named "gausn1.csv" (101.0B) and a "Click here" button to download the CSV with the proposed data for optimisation.

Optimisation of two functions

Global objective

Find θ corresponding to the highest value of y_1 and the lowest value of y_2

Graphical representation



How it works on MADGUI



Navigation

- Main Page
- Prediction
- Bayesian
- About
- Contact

Upload your data

Drag and drop file here
Limit 200MB per file • CSV, XLSX

Browse files

d10.csv
123.0B

Firstly, you will have to upload your data. To do that you have to click on the "Browse File" button in the sidebar.

Your data from the file d10.csv are represented below, you can check if everything is alright.

	Sample	theta	y1	y2
	0	1	-4.7688	0
	1	2	8.6382	3.0252
				-2.1154

2 - Selection of your features and targets for the project

On this section of the program, you must select which columns of your dataset are the features you want to analyse and which columns are the targets that you want to predict or improve.

Columns with a standard deviation of 0 are already deselected, as well as columns that contain text. This is done automatically to eliminate columns that do not provide useful information for analysis and prediction.

Features - Unselected the one you don't need :

theta x

Select your targets then click the submit button :

y1 x y2 x

Submit

The data that you selected (Features + Targets) are :

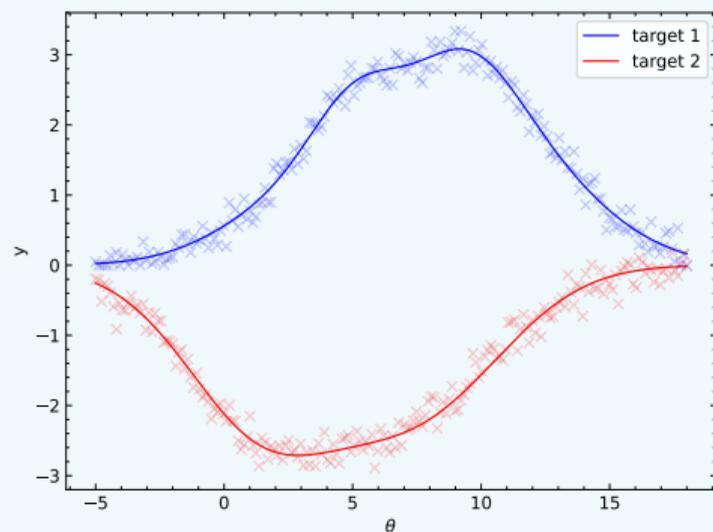
	theta	y1	y2
	0	-4.7688	0
	1	8.6382	3.0252
			-2.1154

Optimisation of two functions

Global objective

Find θ corresponding to the highest value of y_1 and the lowest value of y_2

Graphical representation



How it works on MADGUI

The screenshot shows the MADGUI interface with the following configuration:

- Navigation:** Main Page, Prediction, Bayesian (selected), About, Contact.
- Global objective:** point over the current best observed value in the optimization process. It helps guide the search for the global optimum in an efficient manner.
- How many target do you want to optimise? (max 3):** 2
- Select which target you want to optimise, be careful to not choose the same target twice:** y1
- Select if you want to maximize/minimize this target:** maximize
- Select which target you want to optimise, be careful to not choose the same target twice:** y2
- Select if you want to maximize/minimize this target:** minimize
- Select the importance factor of y1 in pourcentage, compare to y2:** 50
- Select the number of sample that you want:** 3
- Execute the Bayesian Optimisation** button.
- Upload your data:** Drag and drop file here (Limit 200MB per file • CSV, XLSX). Browse files button.
- File upload:** d10.csv (123.0B)
- Data table:**

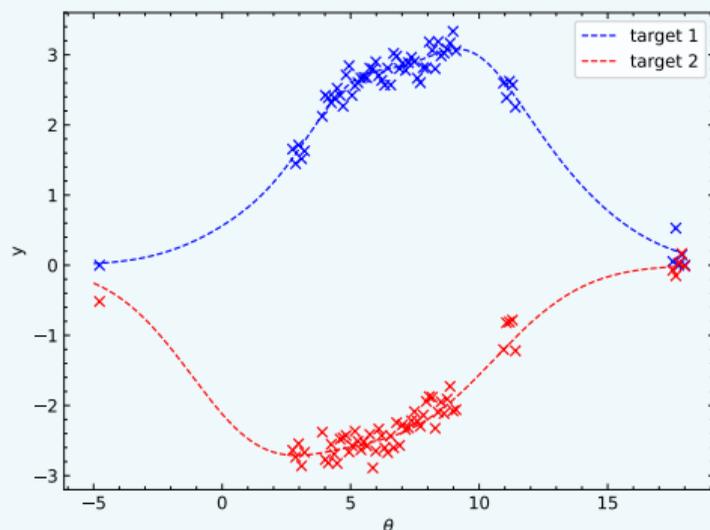
	theta
0	8.9
1	3.1
2	17.9

Optimisation of two functions

Global objective

Find θ corresponding to the highest value of y_1 and the lowest value of y_2

Graphical representation



How it works on MADGUI

The screenshot shows the MADGUI interface with the following configuration:

- Navigation:** Main Page, Prediction, Bayesian (selected), About, Contact.
- Global objective:** point over the current best observed value in the optimization process. It helps guide the search for the global optimum in an efficient manner.
- How many target do you want to optimise? (max 3):** 2
- Select which target you want to optimise, be careful to not choose the same target twice:** y1
- Select if you want to maximize/minimize this target:** maximize
- Select which target you want to optimise, be careful to not choose the same target twice:** y2
- Select if you want to maximize/minimize this target:** minimize
- Select the importance factor of y1 in pourcentage, compare to y2:** 50
- Select the number of sample that you want:** 3
- Execute the Bayesian Optimisation** button.
- Upload your data:** Drag and drop file here (Limit 200MB per file • CSV, XLSX). Browse files button.
- File upload:** d10.csv (123.0B)
- Data table:**

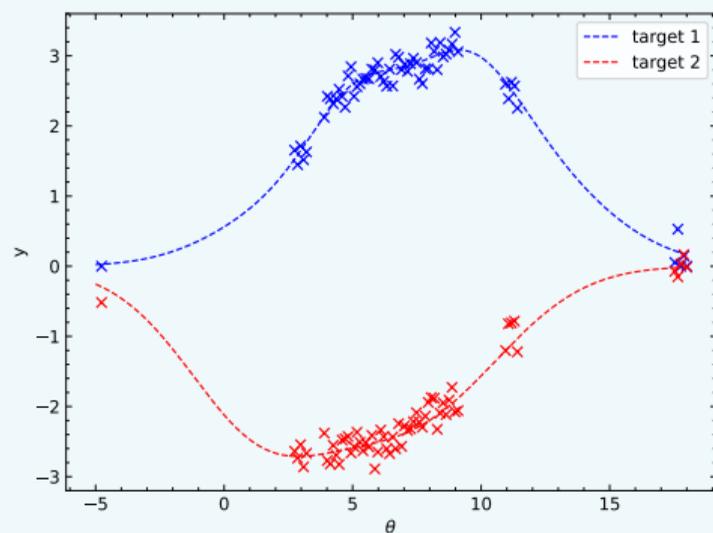
	theta
0	8.9
1	3.1
2	17.9

Optimisation of two functions

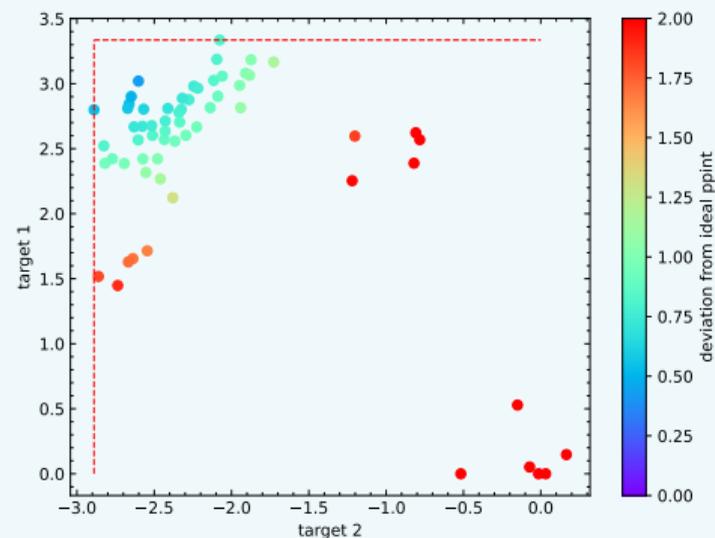
Global objective

Find θ corresponding to the highest value of y_1 and the lowest value of y_2

Graphical representation

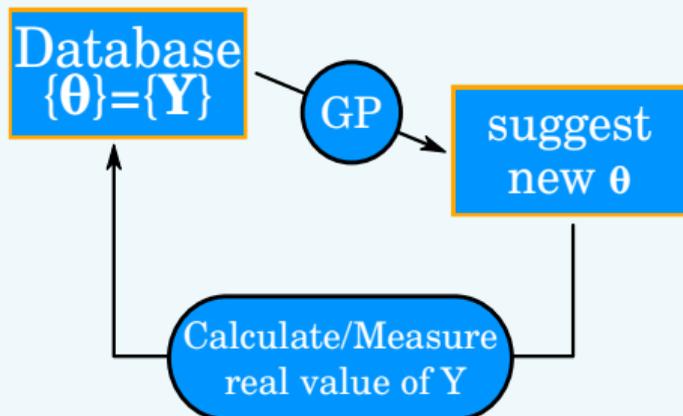


Results representation



Real examples

Framework in precedent examples



problem

Only one compound can be calculated at a time

SPS application

Optimise SPS synthesise parameter to obtain the best zT on thermo-electric compound.

The screenshot shows the MADGUI web interface. On the left is a navigation menu with options: Main Page, Prediction, Bayesian, About, and Contact. The main content area displays a table with 10 samples and their corresponding parameters (HR, Sirt, Stept, CR, UP) and the resulting zT value.

Sample	HR	Sirt	Stept	CR	UP	zT	
0	1	100	773	15	100	50	0.276
1	2	100	973	15	100	50	0.269
2	3	50	773	15	50	50	0.27
3	4	100	773	15	100	50	0.269
4	5	175	773	15	170	50	0.271
5	6	100	773	15	100	50	0.3
6	7	100	750	15	100	50	0.269
7	8	100	873	15	100	50	0.273
8	9	100	873	15	100	60	0.247
9	10	100	773	15	100	60	0.26

2 - Selection of your features and targets for the project

On this section of the program, you must select which columns of your dataset are the features you want to analyse and which columns are the targets that you want to predict or improve.

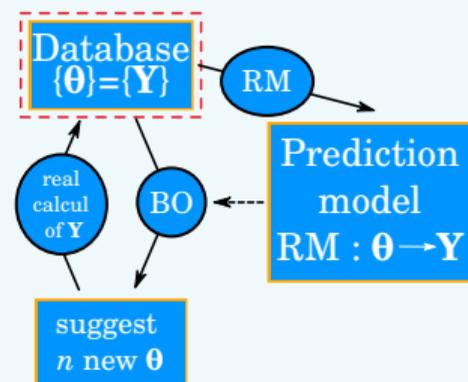
Columns with a standard deviation of 0 are already deselected, as well as columns that contain text. This is done automatically to eliminate columns that do not provide useful information for analysis and prediction.

Features - Unselected the one you don't need :

HR × Sirt × Stept × CR × UP ×

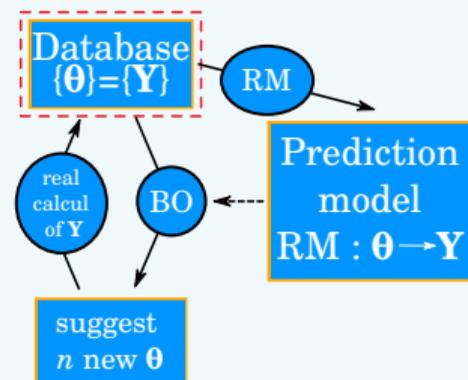
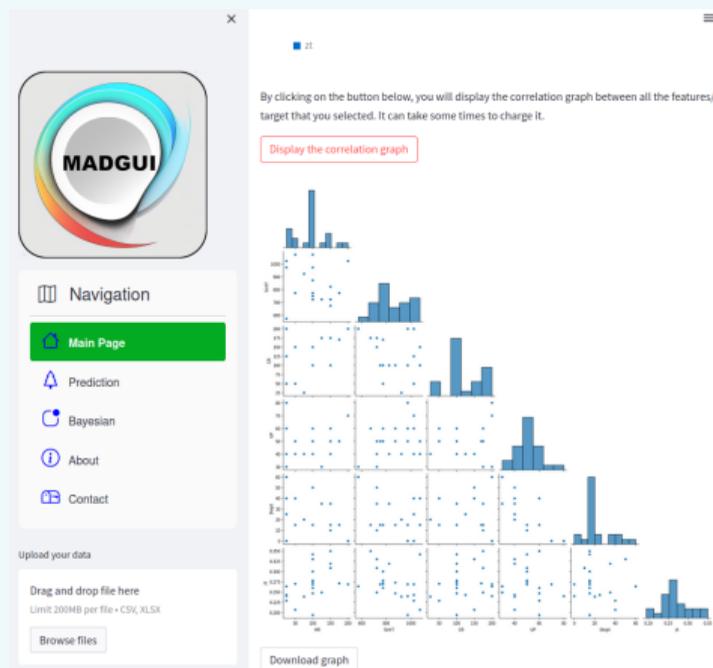
Select your targets then click the submit button :

zT ×



SPS application

MADGUI can display some information on database



SPS application

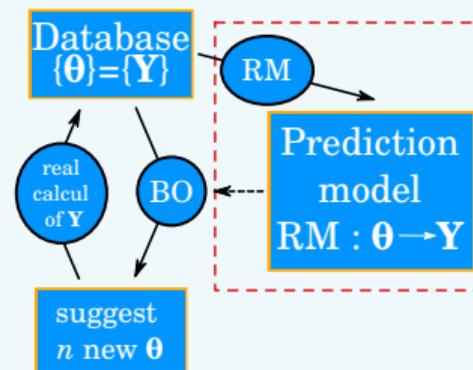
- Models

ElasticNet : very quick
 RandomForestRegressor : usually good
 XGBRegressor : usually good

- Cross validation

LeaveOneOut : for less than 10 data
 K-Fold (3-4-5) : for more data

The screenshot shows the MADGUI web application. On the left is a navigation menu with options: Main Page, Prediction (highlighted in green), Bayesian, About, and Contact. The main content area is titled "Target's Prediction using different methods (ElasticNet, RandomForest and XGBRegressor)". It contains several dropdown menus: "Explanation" (set to "zt"), "Select which target you want to predict" (set to "zt"), "Select which method of prediction you want to use" (set to "RandomForestRegressor"), and "Select which method of cross validation you want to use" (set to "K-Fold"). Below these is a field for "Choose how many subsets do you want to use, it has an impact only if you selected K-fold" with the value "4" and a "Submit" button.



Conclusion

Regression machine learning model

- Allows you to predict information about a compound rather than calculate/measure it
- Need a lot of data

Bayesian Optimisation (active learning)

- Allows you to optimise the properties of a compound with a minimum of samples
- Can be applied to calculated or experimental value

MADGUI

Is a fully graphical interface that makes BO accessible to researchers without extensive programming experience

Link 
IRL 3629

Thanks for your attention

