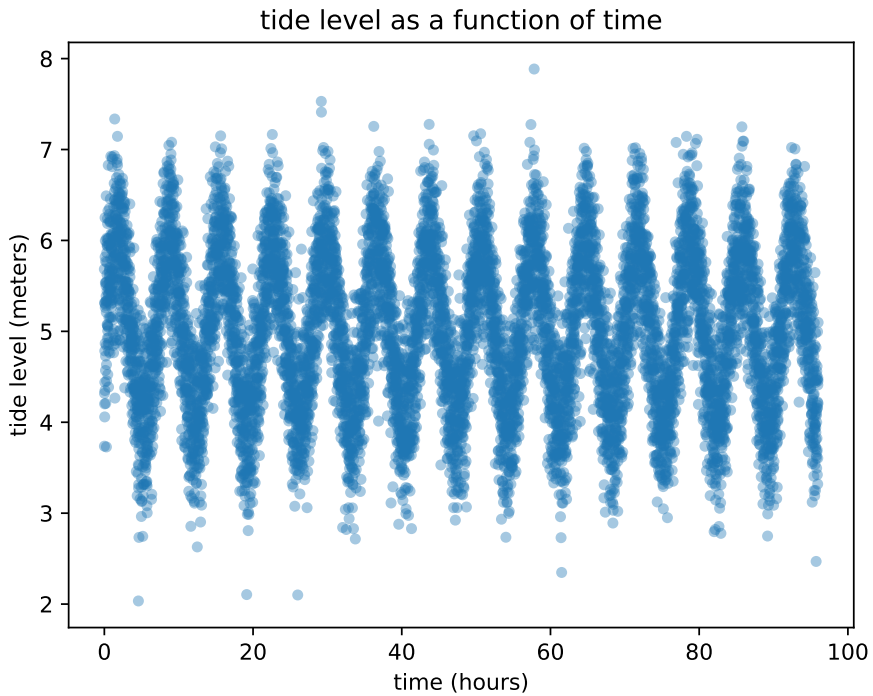
Machine learning I, supervised learning: nonlinear regression

Content

In this example, we will study a **time series** (série temporelle). The dataset contains the tide level (in meters) as a function of the time (in hours).



Tide Level

We have a dataset containing the tide level in meters as a function of time in hours.
Our goal will be to **predict** the tide level as a function of time.

Tide level

Exercice 1 : **Finding a function**

We would like to **model** the tide level as a function f of the time.

Tide level

Exercice 1 : **Finding a function**

We would like to **model** the tide level as a function f of the time.

What would you suggest ?

We could use a sine function. What would the **parameters** be ?

Tide level

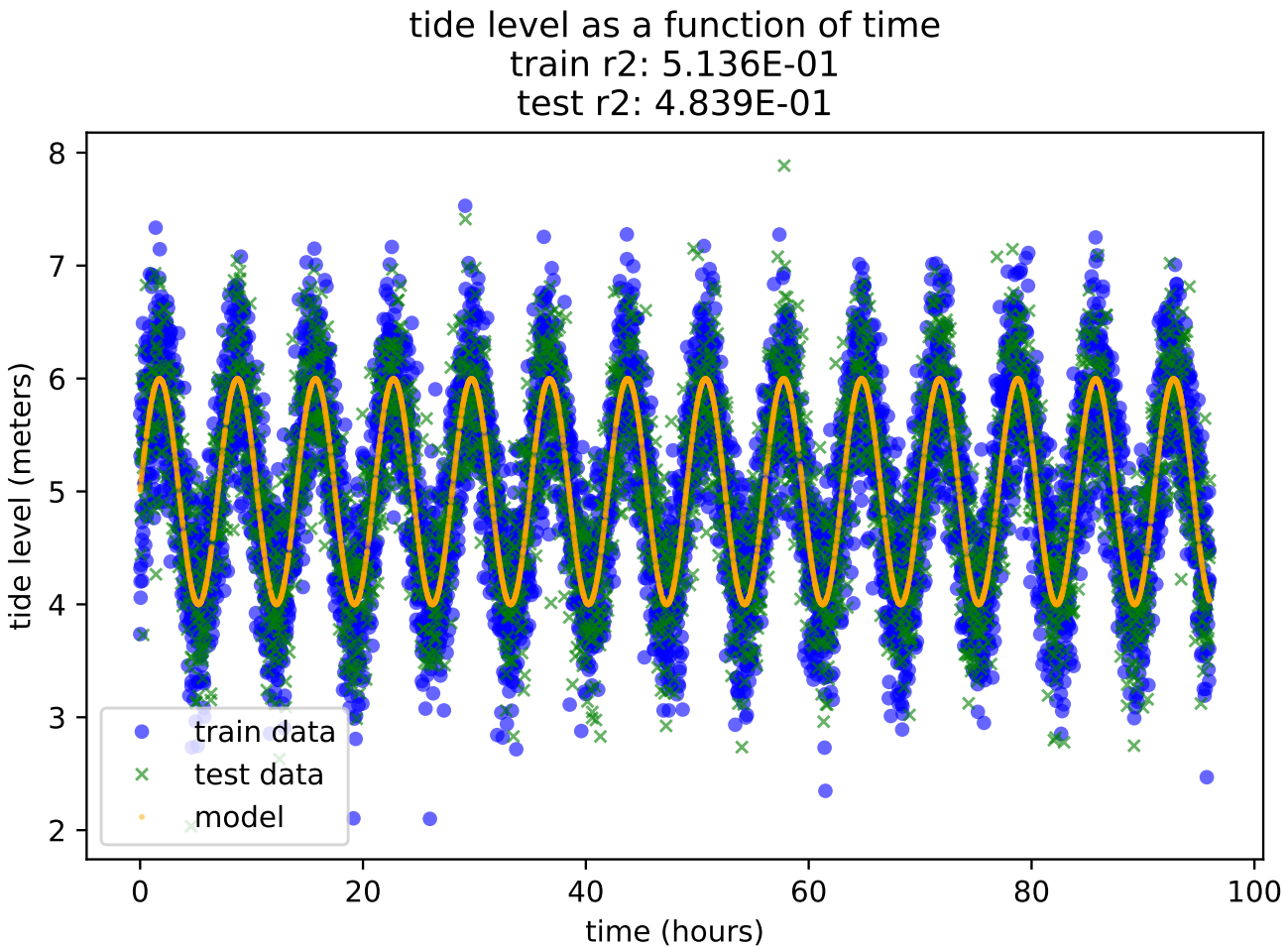
Exercice 1 : **Finding a function** We would like to model the tide level as a function f of the time.
We could use a sine function. The parameters are :

- Amplitude
- pulsation (analog of frequency)
- phase
- offset

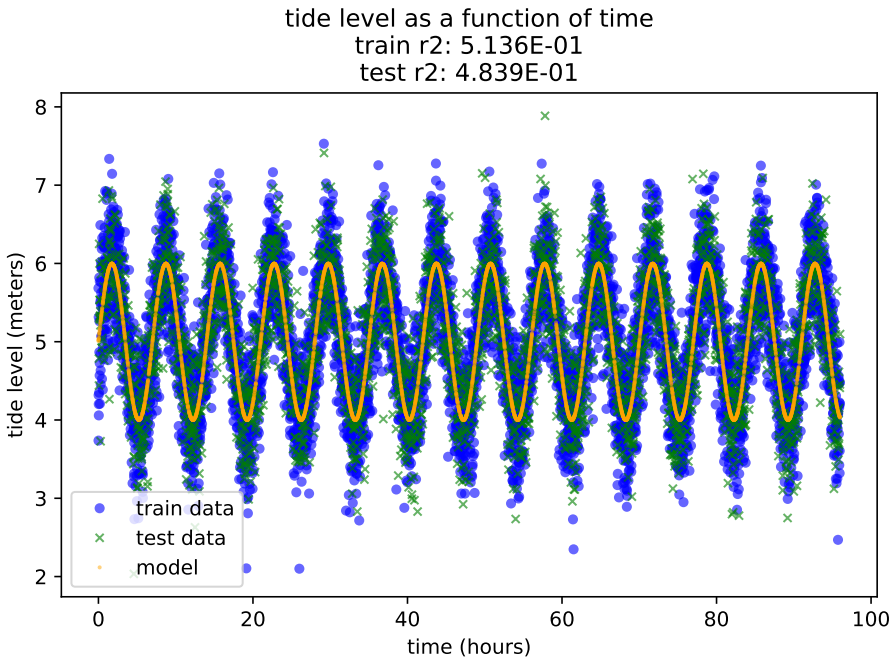
$$\tilde{f}(t) = A\sin(\omega t + \phi) + B \tag{1}$$

In **nonlinear_regression.py**, we have a file **main_learn_estimator.py** that learns a function of the previous form, based on the dataset.

Tide level


tide level as a function of time
train r2: 5.136E-01
test r2: 4.839E-01
tide level (meters)
time (hours)
train data
test data
model
8
7
6
5
4
3
2
0
20
40
60
80
100

Tide level



The inaccuracy comes from the **variance** in the data, which comes from **random noise**, due to the existence of a large number of variables playing a role in the measurements. By constraining the function shape, we avoid overfitting.