Instituto de Radioastronomía y Astrofísica - UNAM

Data Analysis Working Group IRyA

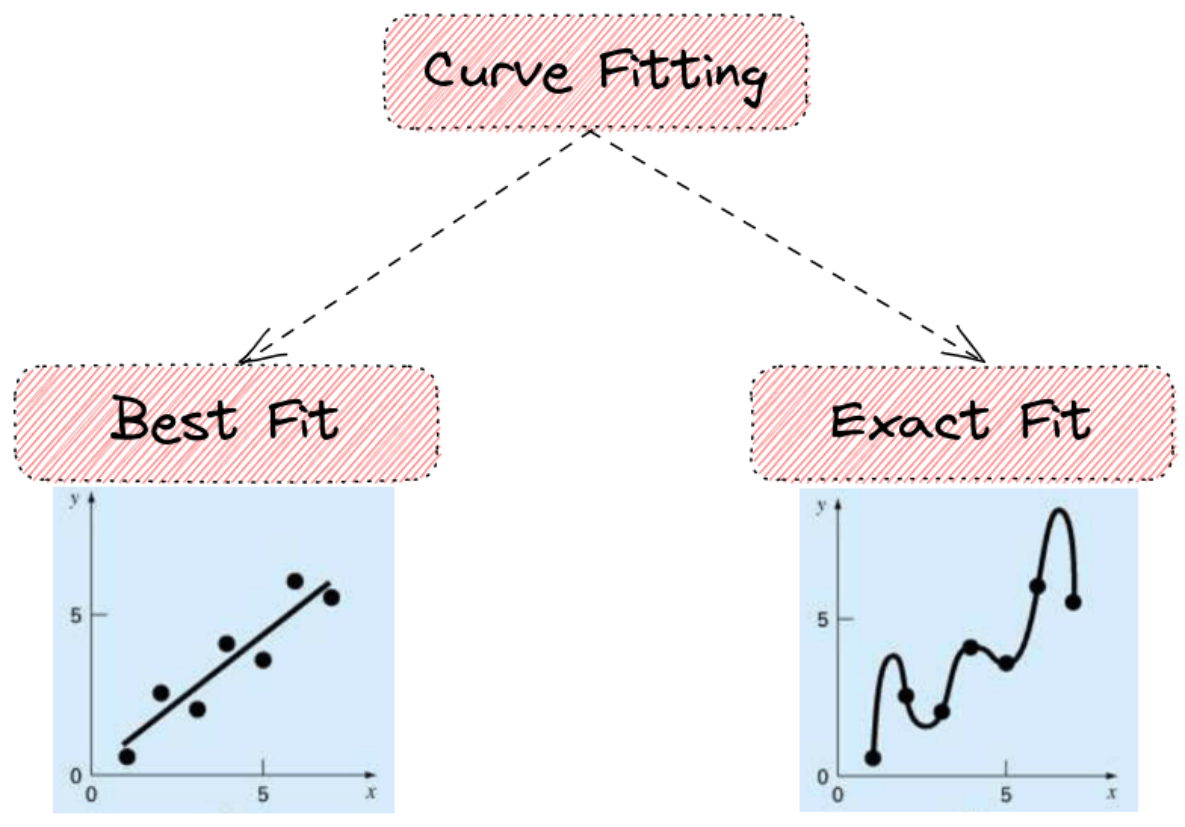
Data fitting with Python

By Omar Ulises Reyes Amador

May 23, 2023

What is fitting?

Curve fitting is a type of optimization that finds an optimal set of parameters for a defined function that best fits a given set of observations.



Data fitting

Python is a power tool for fitting data to any functional form.

You are no longer limited to the simple linear (linear regression) or polynominal functions (polynomial regression) you could fit in a spreadsheet program.

You can also calculate the standard error for any parameter in a functional fit.

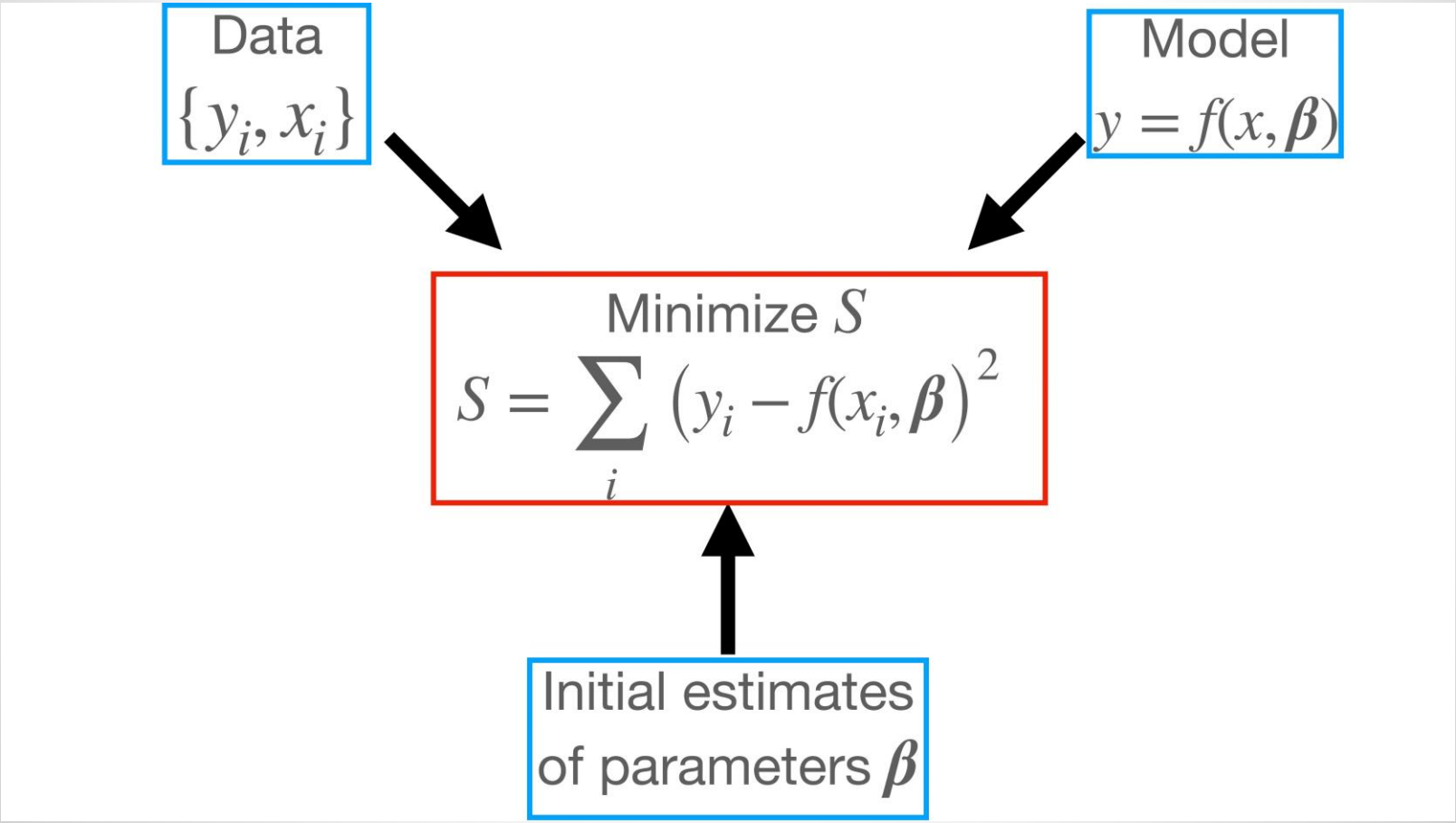
Tools for fitting

Python offers a wide range of tools for fitting mathematical models to data.

1. numpy.polyfit (and others numpy classes) → Linear models
2. scipy.optimize.curve_fit
3. pymodelfit
4. lmfit

} Non-linear models

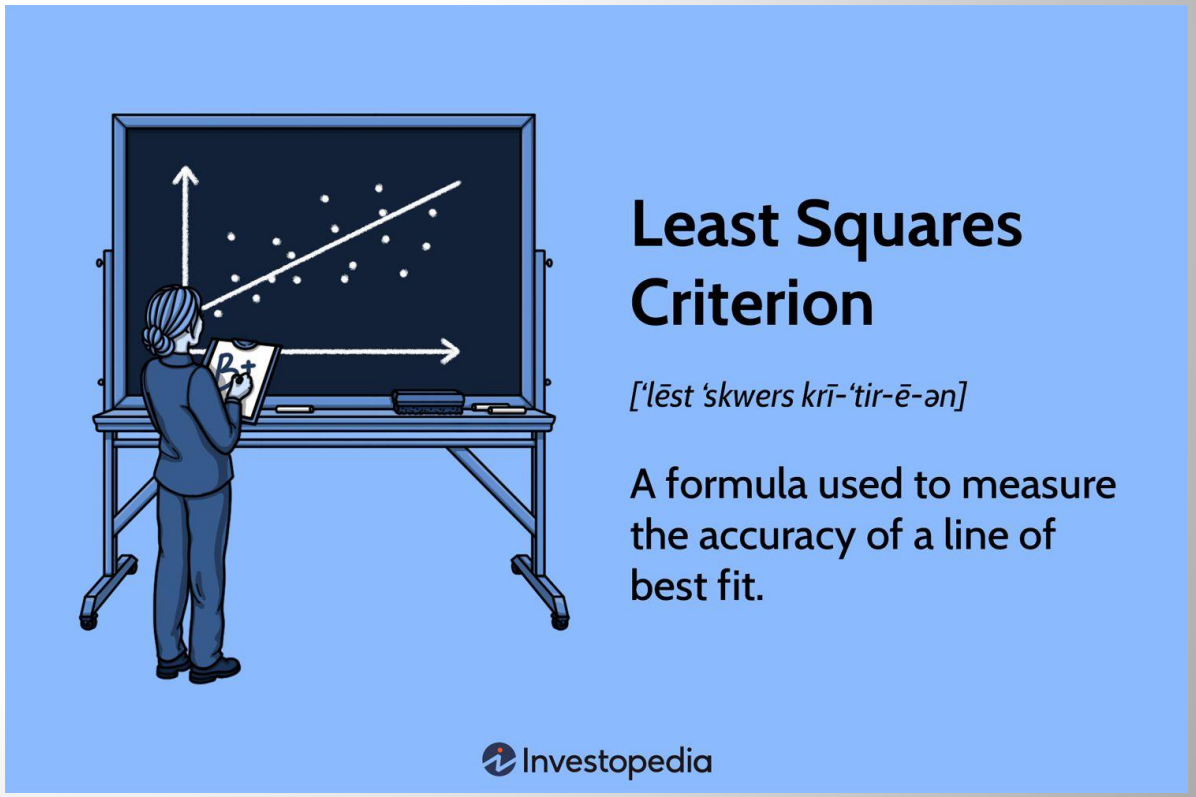
How do they work?



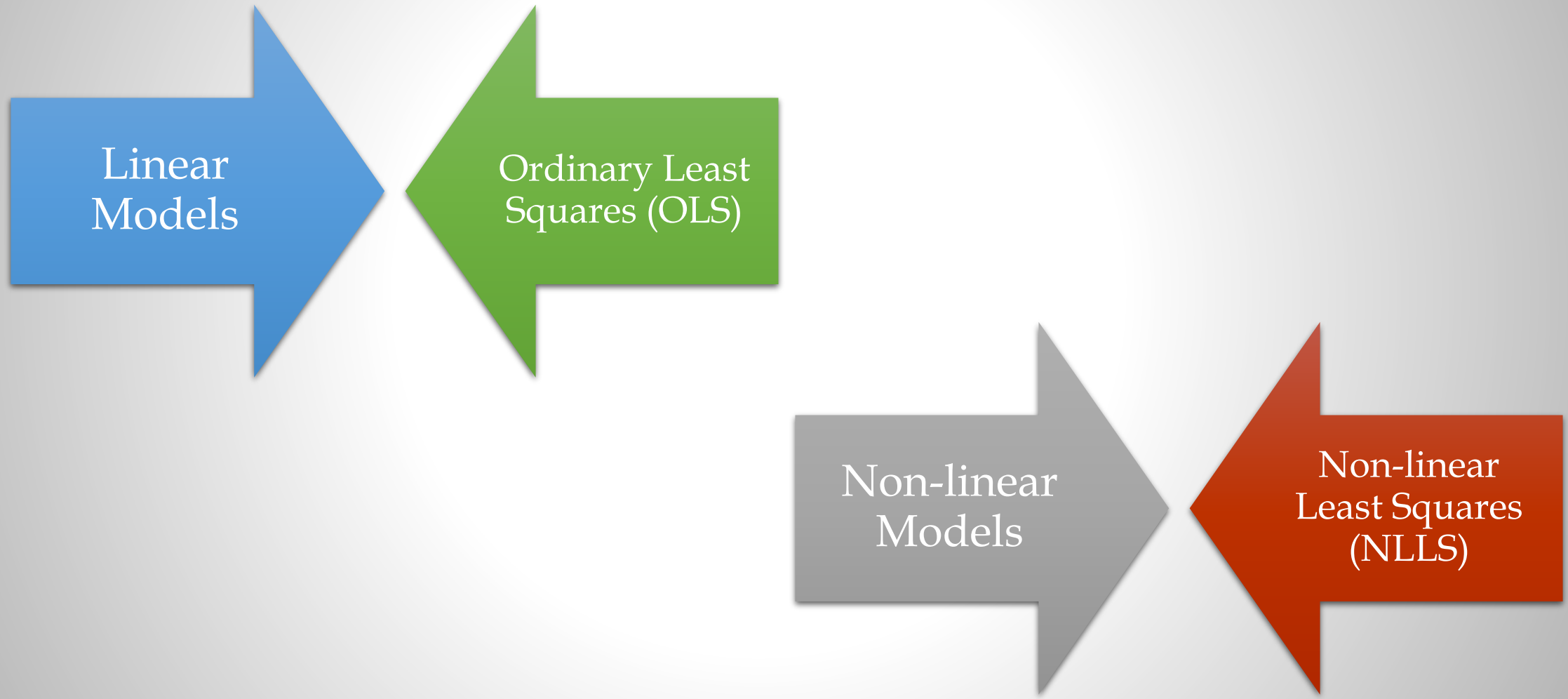
Least squares criterion

$$\mathrm{RSS} = S = \sum_{i=1}^{n} [y_i - f(x_i, \boldsymbol{\beta})]^2 = \sum_{i=1}^{n} r_i^2$$

The objective of any Least Squares method is to find estimates of values of the parameters (β) that minimize the sum (S) of squared residuals.

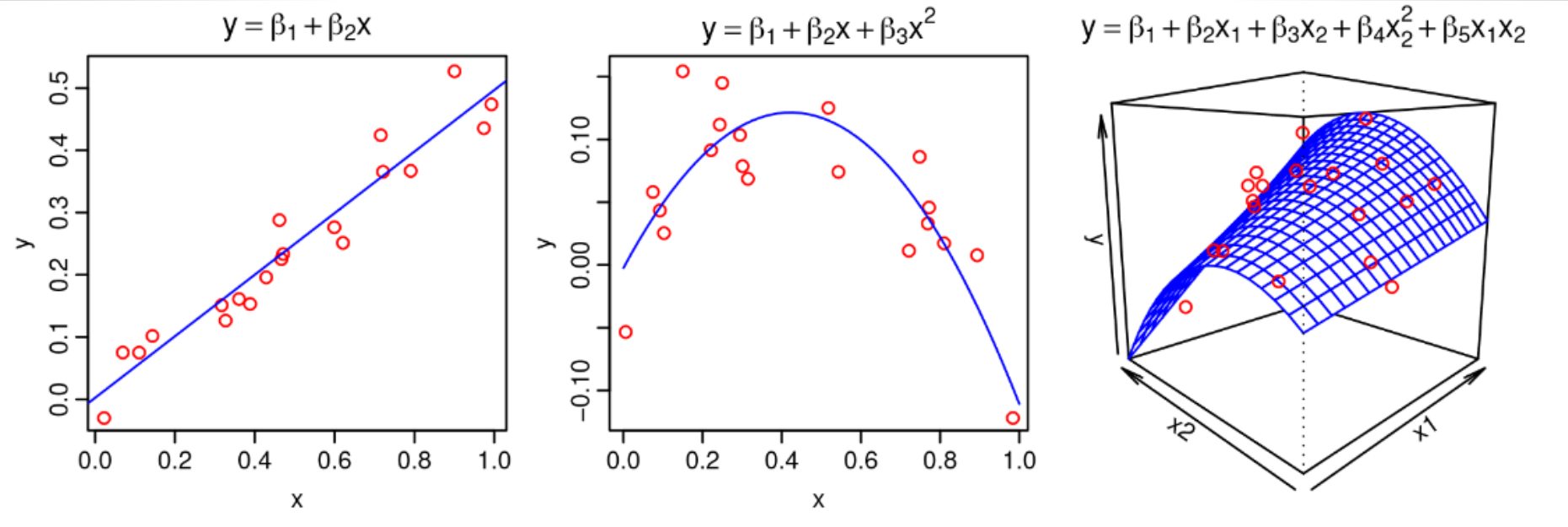


Linear and Non-linear



Linear Models

- Linear models are mathematical functions that can be created with a linear combination of variables and coefficients.
- They are easily fitted using Ordinary Least Squares.
- They can include curved responses (e. g. polynomial regression)

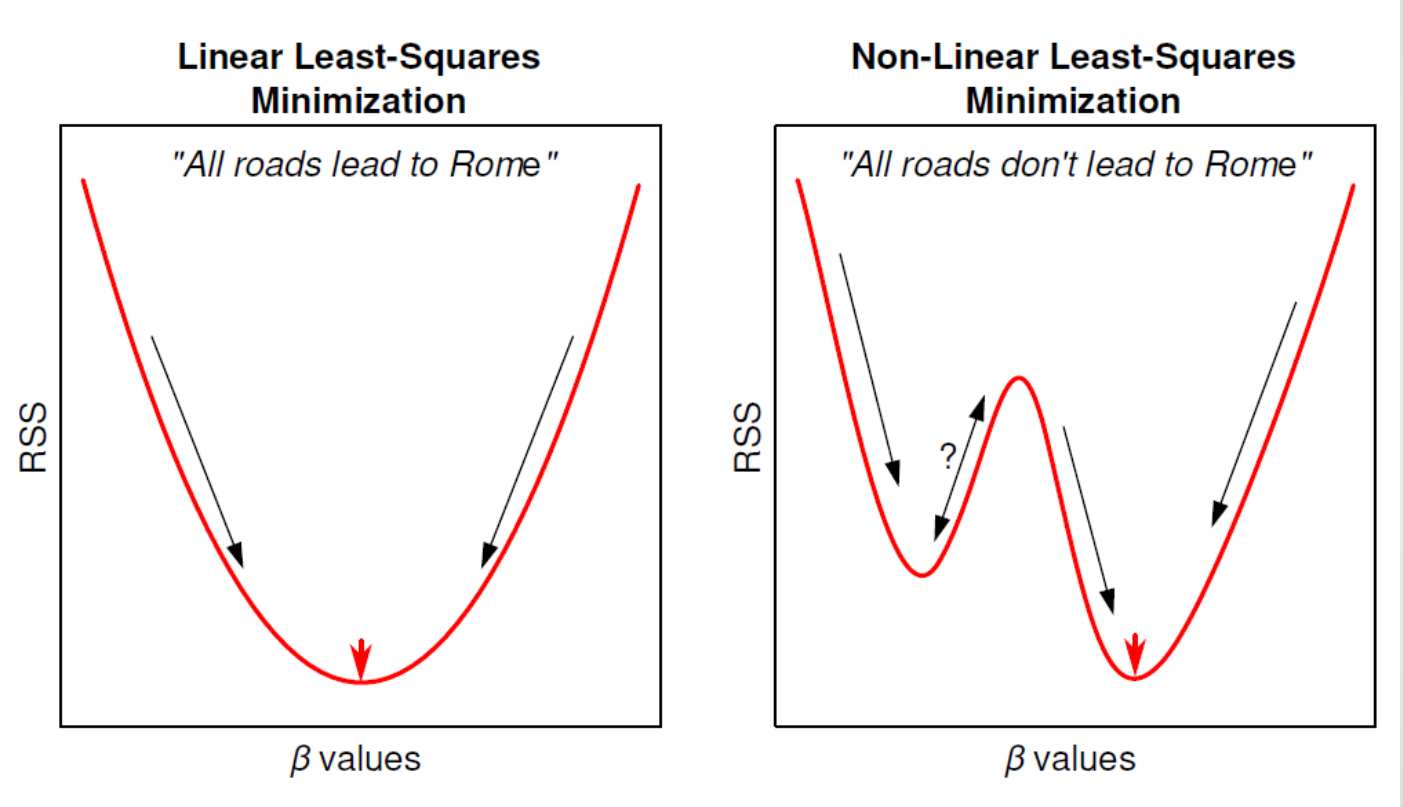


Non-linear Models

- $y_i = \beta_0 x_i^{\beta_1} + \varepsilon_i$
- $y_i = \beta_0 + \beta_1 x_i^{\beta_2} + \varepsilon_i$
- $y_i = \beta_0 e^{\beta_2 x_i} + \varepsilon_i$
- $y_i = \frac{\beta_0 x_i}{\beta_1 + x_i} + \varepsilon_i$

In all of these, at least one term is non-linear.

OLS vs NLLS



If the model is linear, the least-square solution is exact.

If a model is non-linear, an exact least-square solution is impossible.

How does NLLS work?

- We can use a computer to find an approximate but close-to-optimal least-squares solution.
- In general:
  - Choose initial values for the parameters to estimate.
  - Adjust the parameters iteratively such that the RSS is gradually decreased.
  - Eventually, a combination of the parameters that is very close to the desired solution can be found.

scipy.optimize.curve_fit

https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.curve_fit.html

https://education.molssi.org/python-data-analysis/03-data-fitting/index.html

PyModelFit

https://pythonhosted.org/PyModelFit/index.html

https://pythonhosted.org/PyModelFit/over.html#fitting-a-model-to-data

LMfit

https://pypi.org/project/lmfit/

https://mhasoba.github.io/TheMulQuaBio/notebooks/Appendix-NLLS-Python.html