



LABORATORIO DI:

METODI E MODELLI MATEMATICI IN PYTHON

A CURA DI: **ANTONIO MIRARCHI & GIUSEPPE TROTTA**

<https://www.labmetodiemodelli.it/>

2° Test intermedio

Manuale di sopravvivenza al laboratorio



4 esercizi



Argomenti lezioni
dalla 1 alla 6



Consegna via mail entro le ore 17:00
antonio.mirarchi@thinkopen.it

<http://www.labmetodiemodelli.it/>

Dove eravamo rimasti?



Manipolare I dati con
Pandas e Matplotlib

<https://www.labmetodiemodelli.it/>



Let's Code!

<https://www.labmetodiemodelli.it/>

1

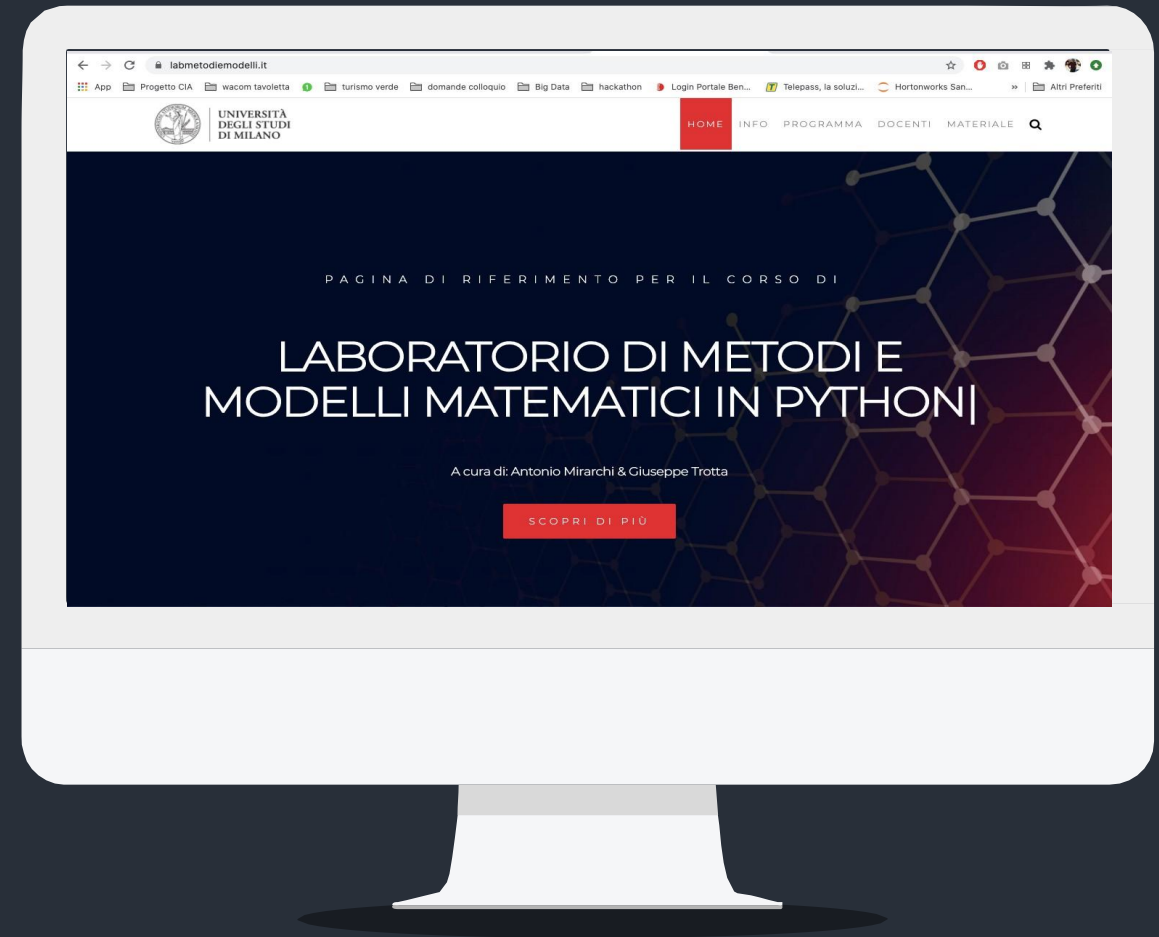
Machine Learning

2

Costruire un Modello Predittivo

3

Regressione



Build a Predictive Model

01

Import Libraries

02

Import Dataset

03

Cleaning Datasets – Null Value, Remove column etc.

04

Pre-visualizations and Correlation

05

Training a model (Train-Test_Split)

Regression attempts to predict one dependent variable (usually denoted by Y) and a series of other changing variables (known as independent variables, usually denoted by X).

Linear Regression is a way of predicting a response Y on the basis of a single predictor variable X . It is assumed that there is approximately a linear relationship between X and Y . Mathematically, we can represent this relationship as:

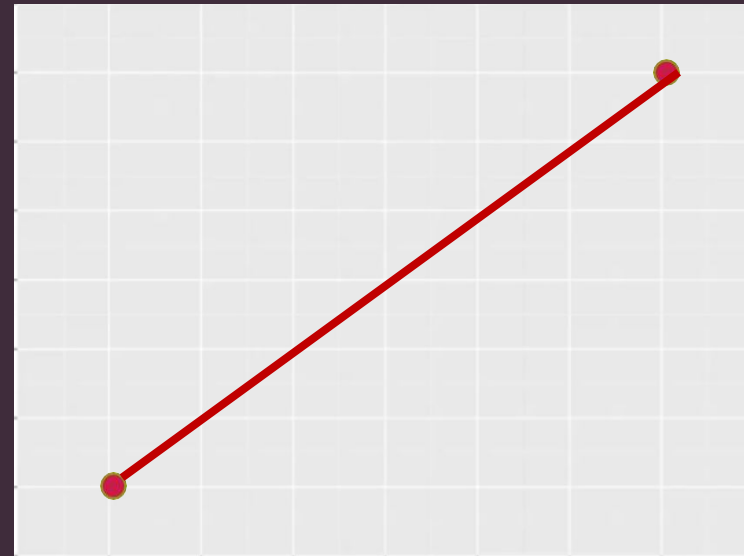
$$Y \approx \alpha + \beta X + \varepsilon$$

where α and β are two unknown constants that represent intercept and slope terms in the linear model and ε is the error in the estimation.

Let's take the simplest possible example. Calculate the regression with only two data points.

Here we have 2 data points represented by two black points. All we are trying to do when we calculate our regression line is draw a line that is as close to every point as possible.

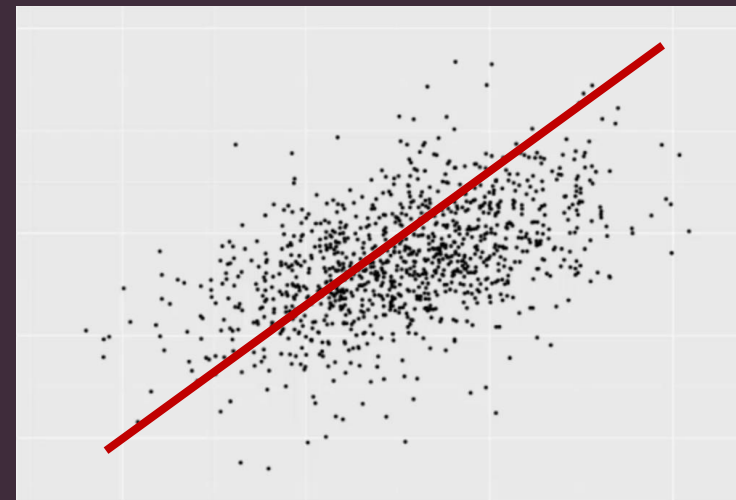
Here, we have a perfectly fitted line because we only have two points.



What's up with multiple points?

By applying linear regression we can take multiple X's and predict the corresponding Y values. This is depicted in the plot ->

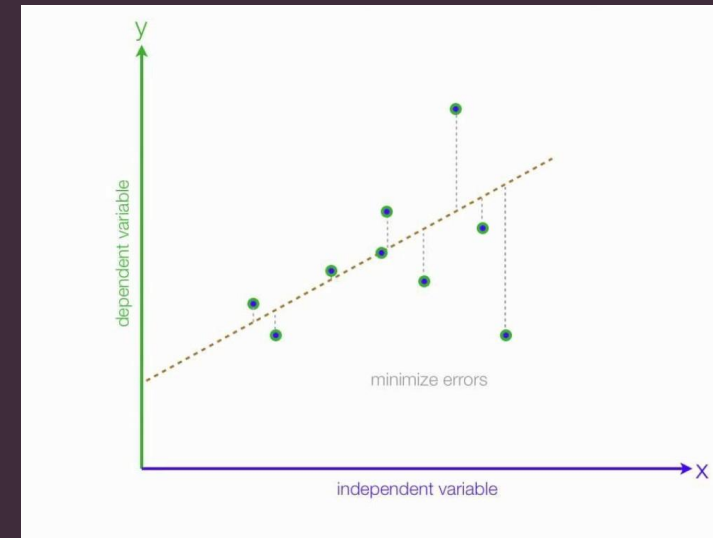
Our goal with linear regression is to minimise the vertical distance between all the data points and our line.

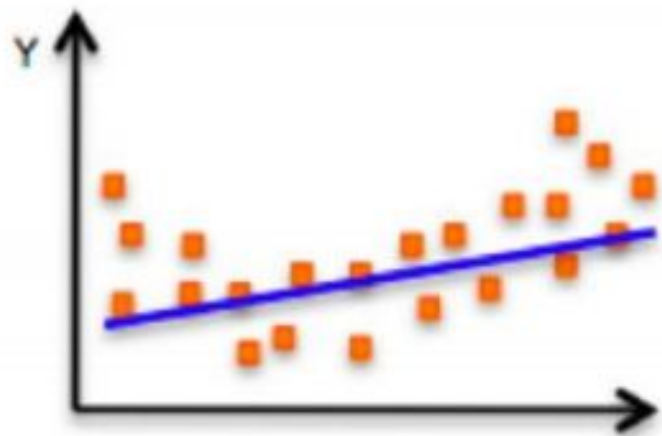


What's up with multiple points?

$$Y = a X + b$$

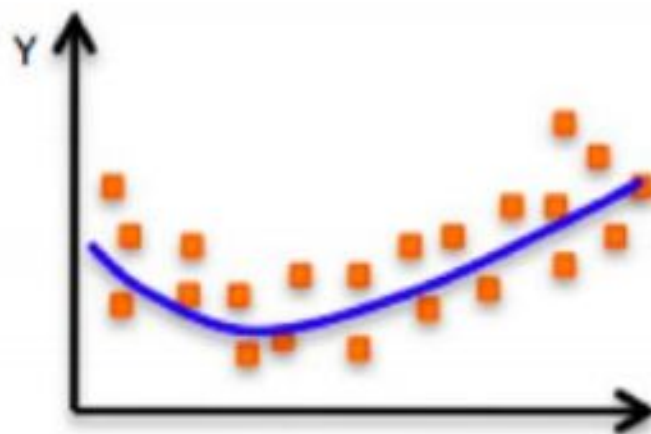
to find the best fit line, which minimizes error (the sum of the square of the distance between points and the line). The distance between the points and line are taken and each of them is squared to get rid of negative values and then the values are summed which gives the error which needs to be minimized.



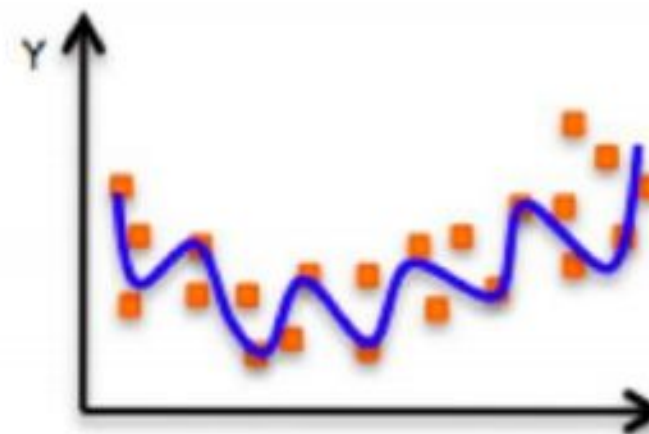


Underfitting

Model does not have capacity to fully learn the data



Ideal fit



Overfitting

Too complex, extra parameters, does not generalize well



Let's Code!

Prossimi Appuntamenti

03

DIC

Costruire Modelli Predittivi – P3

10

DIC

Costruire Modelli Predittivi – P4

17

DIC

Reti Neurali + Ultimo Test Intermedio