

Fundamentos de Data Science

LAB N. 1

In this lab, you will be implementing the polynomial curve fitting examples that we saw in class (from chapter 1 of Bishop's book).

Your dataset will be constituted by n points in 2 dimensions, that originate from the sinusoidal function $y = \sin(2\pi x)$, with the addition of random noise, which is Gaussian distributed with 0 mean and 0.3 standard deviation.

You will implement a program in Matlab to:

- learn the linear coefficients for interpolating polynomials of degree D . (*hint: have a look at the Matlab function `"\`"*)
- output the learned coefficients, so you can look at them
- plot a figure containing:
 - a) The original sinusoidal function in green
 - b) The points as blue circles
 - c) Your interpolating polynomial in red

This simple piece of code will allow you to run experiments to understand the effect of varying the number of training points and the degree of the polynomial.

Today you should aim at finishing a simple script that implements the basic details described above and you should obtain a figure similar to the one I provide below (my code drew this for $n=8$ and $D=3$).

During next week you should try to:

- divide your dataset into training and testing (it will work better for larger values of n);
- measure the error on the training and testing dataset (you can use SSE or RMS);
- write a further piece of code that, given dataset, calculates the error on the training and testing sets for different values of D , and then plots training and testing the error vs D ;
- make your code more modular, factoring it into functions;
- refine your plots (with title, variable names on the axes, etc).

Have fun ! ☺

