Computational Statistics

Exercise sheet 10

To be handed-in for marking until Tutorial of 05/07/2017

Nelson Lima and Luca Amendola, ITP, Heidelberg

05/07/2017

www.thphys.uni-heidelberg.de/~amendola/teaching.html

Problem 1 - Higher-order fitting [10 points]

Start by loading the dataset in file 'rmr_ISwR.dat' provided in the Dropbox folder.

1. Print the data to understand its structure. Then, plot the metabolic rate as a function of body weight to first visualize the data of interest;
2. Following the previous tutorial, perform a first estimation of the linear fit coefficients by finding the slope and the intercept that minimize the residual standard deviation of the fit
$$\sigma = \sqrt{\frac{\Sigma_{\mathrm{i}}^{n}(y_{\mathrm{i}} - y_{\mathrm{fit}})^2}{n-2}} \tag{1}$$
3. Confirm your result by using the function 'lm' and printing its summary;
4. Plot the data and overplot with the linear fits obtained in points 2 and 3. Using the function 'confint', also overplot the 95% confidence limit of the linear fitting found in point 3;
5. Now, using the function 'lm', perform a quadratic fitting to the data. Also do a cubic fitting.
6. Again, plot the data and now overplot with the quadratic and cubic fitting;
7. Lastly, using the linear fit of point 3, predict the metabolic rates for body weights of 150 and 200 kgs, including the errors of the estimations.

Problem 2 - Linear fitting with errors [5 points]

Start by loading the dataset in file 'rmr_ISwR_errors.dat' provided in the Dropbox folder. This should have the same structure of the file used in problem 1 with an added column containing the standard deviation of the measurements of the metabolic rates.

1. Print the data to understand its structure;

2. Compute the linear fit to the data by solving the system of Eqs. (4.16) and (4.17) as described in the most recent version of the lecture notes;

3. Check your results with those of function 'lm', using the errors as weights;

4. Plot the data, using the standard deviations as error bars of the data points, and overplot with the linear fit found before.

Problem 3 - Generalized polynomial fit [5 points]

Imagine that you've got n data points y_i at positions x_i, each with a Gaussian error σ_i. The data points originate from a polynomial model of the form $y(x) = \sum_{\alpha=0}^{m} a_\alpha x^\alpha$, where obviously $n \gg m$. Please

1. formulate a χ^2-functional for the fitting problem,

2. derive a linear system of equations by using the minimum conditions $\partial_\alpha \chi^2 = 0$,

3. write this system in a matrix-vector-notation,

4. formulate the coefficients as *moments including a weighting* with the errors σ_i,

5. invert the linear system for getting the model parameters a_α,

6. solve the special case $m = 0$: What's the solution for a_0?

7. What happens if $n = m$?