ELEC-E9111 - Mathematical computing

Data Import and Fitting

19 September 2024

Please include all your code as `.m` or `.mlx` files in your submission.

1 Linear fit, (5 pts)

First, download the file `data_2022.csv` from MyCourses. The file has data of the trajectory a ball thrown at an observer standing at coordinates $[100, 0]$. The data has increasing amount of noise the further away it is from the observer. In this exercise, you must fit the data in the file using three different models:

1. Linear model with a constant term ($y = a_0 + a_1 x$)
2. Quadratic model ($y = a_0 + a_1 x + a_2 x^2$).
3. Fourth degree polynomial model ($y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4$).

Use your preferred method to calculate the different least-squares fits to the given data. Then, calculate the adjusted R^2_{adj} values of the three models.

$$R^2_{adj} = 1 - (1 - R^2)\frac{n-1}{n-p},$$

where n is the sample size (i.e. the size of y) and p is the number of predictor variables (coefficients) in the model. The unadjusted R^2 value is given by

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2},$$

where y are the measured data values, $\hat{y}$ the values predicted by the model, and $\bar{y}$ is the mean of y. Judging by the R^2_{adj} values, which model seems to give the best fit for the data?

Plot the models in range $x \in [-10 \quad 100]$. Which model do you think gives the best description of a thrown ball? Why the R^2 value is an insufficient statistic for determining which type of model works best?

2 Statistical analysis (5 pts)

a)

Download the file `weather.xls`. It contains weather data gathered by different observers over multiple years. Use analysis of variance (ANOVA) to test if the values recorded by different observers share a common mean value. Run the test for each measured variable. Is there significant difference in values recorded by different observers?

b)

Run the same tests again, but instead of grouping the data by observers, group it by the month of the observation. Is there significant difference in values recorded during different months?

3 Fluorescence lifetime (10 pts)

The file `lifetime.txt` contains measurement data of fluorescence lifetime of Rhodamine 6G.

a)

Import and plot the data. You should see four separate peaks, each peak is a separate measurement.

b)

For this part you will need to extract a segment from the data that corresponds to a single decay and fit the model to that data segment. `findpeaks` with a suitable value for the name-value argument `MinPeakProminence` can be used to find the peak indices. After finding the peaks, remove 10 samples from the end of each section to get rid of the rising edge of the next peak.

Implement a function `[fitresult, gof] = PLDecayFit(time, counts)` that fits an exponential curve to a single decay slope

$$\hat{y} = ae^{-bt} + c,$$

where $\hat{y}$ is the modeled photon count, t is time and a, b, c are the parameters to be fitted. `time` and `counts` are n-by-1 vectors containing the data for a single decay, and `[fitresult, gof]` is the output of the `fit` function.
If you use `cftool` to generate the fitting code, remember to remove the sections defining the excluded points from the autogenerated code. For the final solution the peak locations should be detected from the data instead of hard-coding them into the function.

Hint: To help the fits to converge, you should adjust the single decay time vectors to start from 0 i.e. set for each decay separately `t = t-t(1)`, where `t` is the time vector for a single decay. This will not affect the value of b, but will prevent a from growing too large for the fitting algorithm.

c)

Write a script that reads the data from the input file, fits all four decay slopes, and finds the mean lifetime $\bar{\tau} = 1/\bar{b}$, where $\bar{b}$ is the mean value of the decay constant b from the four fits.