ADS1 exercises – Fitting

1. Fitting. Read the file `india_population.csv`.

    (a) Fit the exponential function
    $$s \cdot \exp(k\, t)$$
    to the population data. `curve_fit` will crash because the python number range is exceeded. Modify your exponential function

    $$s \cdot \exp(k\, (t - t_0))$$

    with t_0 a year in the range 1950 to 2020. You will still not get an acceptable fit because the fit routine ended up in a local minimum. You'll have to find better start values. For finding a good estimate for s consider what the value of $\exp(k \cdot (t - 1950))$ is for 1950. Get an estimate for the annual growth k from the growth column in the dataframe. Calculate and plot function values for your estimate. Vary k until you reach rough agreement between both curves. Use these parameters as start values. Inspect the new fit.

    (b) You will notice a systematic overestimate of the population in recent years. The logistics function descripes exponential growth which is slowing down once saturation sets in. It has the form
    $$\frac{s}{1 + \exp(k\, (t - t_0))}$$
    t_0 is a free parameter. Again you need to get a rough estimate for start values. k again is the growth rate. t_0 is the turning point when the exponential growth is slowing down. Have a guess. A starting guess for s is the population at t_0.

    Plot fit and data. Use the function `err_ranges()` in the attached module errors and the pyplot `fill_between()` function to plot the envelope of the confidence range.

    Use your best fit logistics function to predict population sizes and its uncertainty for 2030, 2040 and 2050.

2. The file `points.csv` contains pairs of x and y values. Read then in and plot them as symbols. You will note a number of outliers. Fit the points and create a new column calculating the fit values and plot them. The fit is distorted by the outliers.

    Calculate the z-scores (do not forget to apply `abs()`). And remove points with z-scores above 3. Check the length of the dataframe. Repeat the fit and calculation of z-scores. Repeat until no further datapoints are removed. Plot the new fit.

As always: follow the PEP-8 recommendations and practise version control.