Data Science Laboratory 1

Lab 8 Practical Challenges

1. Take the height/weight data from height-weight.csv and apply the linear regression model using sklearn's built-in linear model method. Along the way, you should do the following:

   - Print out the first 10 rows

   - Check the column name of the dataset, dimension and whether there are any missing values.

   - Find the correlation between the two variables.

   - Normalise the values using minmax to scale the variables between 0 and 1.

2. Replicate the noisy straight line model as in the notes.

   (a) Iterate brute force over a grid of potential m and c solutions and calculate the χ^2 value for each combination. Visualise this as a 2D image or contour plot. Play around with the level of noise and the number of points.

   (b) Identify the $\Delta\chi^2/\nu = 1$ region and use it to estimate the confidence bounds of the best fit (minimum χ^2/ν) parameters. How do these values compare to the uncertainties obtained through `curve_fit()`?

   (c) Code up by hand the MLE optimization for the same noisy data and model to find the best fit parameters. Use the likelihood surface to estimate the 1σ confidence range of the most likely parameters. Confirm your results agree with the methods above.

   Extension: update the model data to a quadratic and apply MLE to find the best fit parameters.

Remember: use GitHub to maintain your code!