1. Please submit via the LMS portal by 23:59pm on April 25.
2. In submitting your work, you are consenting that it may be copied and transmitted by the University for the detection of plagiarism. Please start with the following statement of originality, which must be signed and dated by all group members: `"This is our own work. We have not copied any of it from anyone else."`
3. This project is worth 20% of your final mark.

In this project you will be investigating COVID-19 data using the least squares method.

COVID-19 data can be downloaded from one of the following sources:
`https://data.europa.eu/data/datasets/covid-19-coronavirus-data-weekly-from-17-december-`
`locale=en`
`https://github.com/CSSEGISandData/COVID-19`
`https://datahub.io/core/covid-19`
`https://github.com/datasets/covid-19`

You can use the `import tool` to read the data into Matlab,
`https://au.mathworks.com/help/matlab/ref/importtool.html`

Two mathematical growth models are:

- **Exponential growth** is described by the differential equation

$$\frac{\mathrm{d}P}{\mathrm{d}t} = gP, \qquad P(0) = p.$$

  Here $P(t)$ is the Population size at time $t \geq 0$, p is the initial population at $t = 0$, and $g > 0$ is the growth parameter for the population. The solution is given by

$$P(t) = p \exp(gt).$$

- **The Verhulst logistic growth** equation is given by

$$\frac{\mathrm{d}P}{\mathrm{d}t} = gP\left(1 - \frac{P}{c}\right), \qquad P(0) = p.$$

  The quantities P and g, p are the same as in the exponential model. The extra parameter c is the carrying capacity. This differential equation can be solved using separation of variables and partial fractions, and the solution is given by

$$P(t) = \frac{pc}{p + (c - p)\exp(-gt)}.$$

In our application the population P is the number of COVID-19 infections. As you cannot be sure how many people were infected at $t = 0$ (and in fact you do not know when infection started), you need to obtain a value of p by fitting the entirety of the data. Thus, the optimisation problem for exponential resp. logistic growth is 2- resp. 3-dimensional.

This is *your* investigation. However, here are some ideas:

- Either focus on one country, or do a comparison between several countries.
- Connect the “flattening of the curve” to social distancing measures/superspreader events at specific times $t_1, t_2, \ldots$.
- Compare the mathematical models, or use different models for different time-intervals.
- Consider other mathematical models.
- Compare different optimisation algorithms, such as the gradient or steepest descend method, Newton’s method or the downhill simplex method.
- Focus on a method we have not covered in this subject, e.g. Levenberg-Marquardt, cf.

  `https://courses.seas.harvard.edu/courses/am205/fall13/AM205_unit_1_chapter_4.pdf`
- Explore and utilise in-house Matlab optimisation routines.
- Compare data sets/sources.
- Find your own niche.

Finally, your report (1 report for each project group) must be written in the style of a scientific or technical research report. Guidelines are given in the documents *How to write the report* and *Project rubric*, available in the section Assessment on LMS.