CSCI 2202 Computer Modelling for Scientists
Lab C Due: 2 Nov 2021

- **Reading Data Files:**
- **Fitting Regression Lines to data**

(1) Write a function `readDatFile(inFile)` to read in the coffee cooling data. The function should take a *filename* as a parameter (input by the user) and should return the `numpy arrays` x, y containing the data.

(2) **Using Regression to Determine Model Parameters:**

(a) The file `coffeeCooling.txt` alongside, holds the `time, Temp` data for coffee cooling. Create a program that reads the data from the file and uses `matplotlib.pyplot` to scatter-plot the data. Use the `pyplot` functions `xlabel( ), ylabel( )` to label the axes appropriately.

(b) A model of coffee cooling in a room at $T_0 = 22^oC$ is: $\Delta T = T_{n+1} - T_n = -k(T_n - T_0)$. Use the function `scipy.stats.linregress(x,y)` (usage: `slope, intercept, r_value, p_value, std_err = stats.linregress(x,y)`), to make a fit of the coffee data given above. Make a plot of the regression line fit, superposed on a scatter plot of ΔT_n v/s $T_n - T_0$.
Note: The `r_value` *gives the correlation coefficient of the fit - the closer the value is to +/- 1, the better. The* `p_value` *gives the validity of the model. It is the result of testing if the slope of the fit is different from zero. The smaller the* `p_value` *the better.*

(3) **Fitting a Power Law Function** Small nanoparticles of soot suspended in water start to aggregate when salt is added. The average radius r of the aggregates is predicted to grow as a power law in time t according to the equation $r = r_0 t^n$. (The data is in the file `sootAggregation.txt`. The columns are: *time (mins)*, *Avg. Radius (nm)*, *Uncertainty in Radius (nm)*). Taking the logarithm of this equation gives $\ln r = n \ln t + \ln r_0$. Th us the data should fall on a straight line if $\ln r$ is plotted vs $\ln t$. Fit a regression line to the data using $\ln r = n \ln t + \ln r_0$ and find values for n and r_0.

(4) `yeastGrowth.txt` contains the biomass of yeast (as a function of times in hours), that is growing in a *resource limited* environment.

(a) Use `readDatFile(inFile)` to read and scatter-plot the data. Note that the growth of the yeast curves and saturates as it reaches the *carrying capacity* of the yeasts' environment. We will use the carrying capacity of the yeasts' environment to be: $p_m = 665$.

(b) The growth of the yeast if modelled as:

$$\frac{dp}{dt} = r(p_m - p) \cdot p$$

In addition we know $p(0)$, the initial amount of yeast. The equation can be solved exactly to give:

$$\ln \frac{p}{(p_m - p)} = r \cdot p_m \cdot t + C$$

Here we C is an arbitrary constant of integration (you will find this) Create a plot to test the model against the data. Using the `yeastGrowth.txt` data, plot $\ln \frac{p}{(p_m - p)}$ v/s t

(use `matplotlib.pyplot` function `semilogy(t, z)` to plot $\ln z$ v/s t.

(c) Make a regression fit to the model above and find the slope ($r \cdot p_m$) and intercept (C) of the line. (Note: we know $p_m = 665$ from the data).

(d) The solution can be written in another form:

$$p(t) = \frac{p_m}{1 + e^{-r*p_m(t-t^*)}}$$

where t^* is the time that $p(t^*) = p_m/2$, which, gives: $t^* = \frac{-C}{r * p_m}$. Find this value and make a plot of $p(t)$ v/s t. This is known as logistic growth model, invented by Pierre-Francois Verhulst.