BIOE 498/599 Midterm: Feb 2021: Sauro

All Python code should be fully documented. These can be provided as python scripts or jupyter notebooks.

This midterm contains 4 questions, with a total of **100** points.

1. (0 points) What is your name?

2. The following function has a number of minima:

$$x^4 + 3x^3 - 9x^2 - 23x - 12 = 0$$

   (a) (15 points) Write your own simulated annealing algorithm that you can use to find the global minimum for this function.

   (b) (20 points) Compare the performance of your own code to the `basinhopping` method from scipy.optimize. Simulated annealing like algorithms don't always succeed at finding the global minimum. Run your code and the basinhopping code to estimate how many times each of the algorithms succeed in finding the global minimum rather than a local minimum? Make sure you report on the mean number of successful rune along with a suitable measure of dispersion of the results.

   Present your comparison in tabular form.

3. Consider the following model:

```
S1 -> S2; k1*S1
S2 -> S3; k2*S2/(Km + S2)
S3 -> S4; k3*S3

S1 = 10;
k1 = 0.2; k2 = 0.1;
Km = 0.4; k3 = 0.03
```

   (a) (10 points) Write a python script that generates synthetic data time course data for this model

(b) (15 points) Fit the model to the synthetic data using the package `lmfit`. You can use all four variables, S_1 to S_4 to fit the model.

Experimentally it might not be possible to measure all four species. If you had to give advice to an experimentalist, which two species might you recommend being the most important to measure? What evidence can you provide to support your recommendation?

(c) (20 points) Investigate the performance of the following optimizers available to lmfit (`https://lmfit.github.io/lmfit-py/fitting.html#the-minimize-function`):

a) Levenberg-Marquardt
b) Nelder and Mead
c) Brute
d) Basinhopping
e) Dual_annealing
f) Differential evolution

Investigate the following performance metrics:

a) How many times was the fit successful for a given method

b) On average how long did it take to fit the model for a given method

c) How well did the methods perform based on the degree of noise in the synthetic data.

Present your results in tabular form.

4. (15 points) Diseases affecting hemoglobin synthesis and function are extremely common worldwide. More than 1000 naturally occurring human hemoglobin variants with single amino acid substitutions throughout the molecule have been discovered. These variants alter hemoglobin structure and biochemical properties. The following table contains data for the mid-regions of the oxygen binding curves of three hemoglobins measured under identical conditions. The oxygen concentration id reported as the partial pressure of oxygen pO2. The three hemoglobins are A (normal form), Hiroshima (His 145 to Asp) and Kansas (Asp 102 to Thr).

Using the data in the table estimate the Hill coefficient for each hemoglobin.

A		Hiroshima		Kansas	
pO2	Y	pO2	Y	pO2	Y
0.67	0.044	0.133	0.045	1.33	0.193
1.33	0.243	0.267	0.179	2.67	0.371
2.00	0.5	0.667	0.62	4.0	0.5
2.67	0.691	1.33	0.822	5.33	0.592
4.00	0.874	2.0	0.948	6.67	0.66