Homework 1
BME 599.17

Submit your responses on Gradescope by Friday, September 26, 2025 by midnight.

Part 1

Consider the following observations of a function $y = g(x_1, x_2)$.

x_1	x_2	y
0.12	0.55	4.2
0.80	0.45	2.2
0.17	0.23	2.8
0.52	0.38	5.0
0.51	0.62	5.5

Using the Kriging equations for conditioning a Gaussian Process, estimate the mean and variance for $g(0.42, 0.33)$ and $g(0.95, 0.15)$.

Part 2

Use a GPR software library to re-solve Part 2. Some recommendations are:

- **Julia**: `GaussianProcesses.jl`
- **Python**: `https://scikit-learn.org/stable/modules/gaussian_process.html`
- **R**: `laGP` package (`https://bookdown.org/rbg/surrogates/chap5.html#chap5library`)
- **Matlab**: `https://www.mathworks.com/help/stats/gaussian-process-regression-models.html`

Part 3

Multiwell plates are a fixture of automated science. The file `GlucoseGrowthData.csv` contains the results of an experiment in a 384 well plate that examined the effects of glucose concentration on total biomass produced by a microbe. The columns in the dataset relevant to this assignment are:

- `Col` and `Row` are the column and row of the well.
- `Final_Glucose_Conc` is the concentration of glucose in the well, in mM.
- `MaxOD` is the maximum optical density (biomass) observed during growth.

1. Plot the maximum OD vs glucose concentration for all wells. Note that the points cluster into two "curves" due to spatial effects in the plate. For example, wells along the edge may heat better, receive more oxygen, and evaporate more quickly during incubation.
2. To remove spatial effects, train two GP models:
   (a) $B(r, c)$ predicts the maximum OD of a "blank" well (with zero glucose) in row r and column c. This model is trained using only the points without glucose.
   (b) Similarly, $G(r, c)$ predicts the maximum OD of the a well with the experiment's maximum glucose concentration.
3. Plot the predicted B and G as a function of location (row and column) in the plate. Do the blanks or the maximum concentration wells display greater spatial variation? Does this make sense?

4. Correct and normalize every well in the dataset. Let $D(r, c)$ be the original data; then the normalized data $\hat{D}$ are

$$\hat{D}(r, c) = \frac{D(r, c) - B(r, c)}{G(r, c)} \quad \forall\, r, c.$$

5. Recreate the maximum OD vs. glucose plot using $\hat{D}$ instead of D. Did you remove the spatial effects?