Université d'Orléans **2025/2026**
Master Physique Fondamentale et Applications (PFA)

Applied machine learning: Regression
Problem sheet 2

This problem sheet focuses on regression analysis using a dataset containing density values against temperature for a common liquid crystal (MBBA). The dataset includes three columns: temperature, density, and error on the density values. Note that there is a phase transition at approximately 45.6°C, indicated by an anomaly in the data.

1. Linear Regression

1. Load the dataset and create a scatter plot of density vs. temperature.
2. Using the analytical formula for linear regression, perform a simple linear regression on the data.
3. Plot the resulting linear fit along with the original data points.
4. Calculate the Mean Squared Error (MSE) for this fit.

2. Non-linear Regression

1. Based on the observed data, suggest a non-linear function that might better describe the relationship between temperature and density.
2. Implement a gradient descent algorithm to fit the non-linear function. Include the following steps in your implementation:
   - Initialize the parameters of your function.
   - Define a cost function (e.g., Mean Squared Error).
   - Implement the gradient descent update rule.
   - Choose an appropriate learning rate and convergence criteria.
3. Plot the fitted non-linear function along with the original data points.
4. Compare the MSE of this fit with those from the linear regression.

3. Heteroscedasticity-aware Fitting

1. Modify your gradient descent algorithm to account for the heteroscedasticity in the data by incorporating the error values provided in the third column of the dataset.
2. Perform the fitting process again using this modified algorithm.