# Testing Regressions & Loss Functions

December 2025

## General Instructions

- **Dataset:** Use the provided 3D spatial dataset (`spatial_data.csv`).
- **LLM Policy:** Usage of LLMs (e.g., Gemini) is allowed **only** for generating plotting code or recalling mathematical formulas. No conceptual derivations or explanation prompts are permitted. All prompts must be kept in a single chat thread.
- **Submission:** Submit a PDF containing your answers and plots, along with the source `.ipynb` file.

### Part A: Spatial Data Exploration

1. Load the dataset and visualize the points in a 3D coordinate system.
2. Describe the observed shape. Comment on its symmetry and the distribution of points across the $X$, $Y$, and $Z$ axes.

### Part B: Global Boundary Estimation

1. Formulate a regression problem to estimate the boundary of the given points. You must treat this as a minimization problem where the loss function is the **Mean Squared Error (MSE)**.
2. Define your hypothesis for the surface equation $f(x, y, z, \theta) = 0$ and use `scipy.optimize.minimize` to find the optimal parameter vector $\theta$.
3. Plot the original data and overlay the estimated boundary surface.

### Part C: Regression on Censored Data

1. **Data Pre-processing:** Filter the dataset by removing all points for which the values of $x$, $y$, **and** $z$ are simultaneously less than zero (i.e., remove the points in the negative octant).
2. Perform the regression analysis again using MSE on this reduced dataset.
3. Compare the resulting boundary with the one obtained in Part B. Discuss how the removal of localized data affects the global fit of the MSE-based model.

## Part D: Robust Regression and Alternative Loss Functions

In engineering applications, data is often incomplete or biased. In Part C, you likely observed that MSE is sensitive to the "missing" information. We now introduce the **Huber Loss** function, which is designed to be more robust. We define residual $\delta$ as the difference between the predicted value and the ground truth. The Huber loss $L_\delta(a)$ for a residual $a$ is defined as:

$$L_\delta(a) = \begin{cases} \frac{1}{2}a^2 & \text{for } |a| \leq \delta \\ \delta(|a| - \frac{1}{2}\delta) & \text{for } |a| > \delta \end{cases}$$

1. Implement the Huber Loss function in Python.
2. Perform regression on the **censored dataset from Part C** using this robust loss function. Use a threshold value of $\delta = 1.0$.
3. Plot the resulting boundary. Contrast this result with the MSE fit from Part C..

*– End of Paper –*