Assignment 6: 3D Classification and Feature Significance

1 Introduction

In this final project, you are provided with a black-box spatial dataset dataset.csv. Your goal is to use Logistic Regression to uncover the hidden geometry of the classification boundary and determine which spatial features are actually driving the model.

2 Part A: Exploratory Data Analysis

1. Load the dataset and generate an interactive 3D scatter plot.
2. By rotating the plot, identify if there is any specific axis along which the data appears to be invariant (unchanging).

3 Part B: Multivariate Logistic Regression

We will assume a quadratic relationship in the horizontal plane while maintaining a linear term for height.

1. Formulate the logit as:

$$z_{logit} = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 z + \beta_4 x^2 + \beta_5 y^2 + \beta_6 z^2$$

2. Minimize the Binary Cross-Entropy loss to find the optimal coefficients.
3. Tabulate your results for $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$.

4 Part C: Discussion on Feature Relevance

1. Look at the optimized value for β_3, β_6 (the coefficient for z and z^2). Is it significantly different from zero?
2. **Question:** Why does the model assign such a value to z? What does this imply about the influence of "height" on this specific classification task?
3. Identify the hidden geometric shape by analyzing the remaining non-zero coefficients.

End of Assignment Series for Logistic Regression.