Assignment 9: RBF Kernels and the 3D Error Surface

Data Science and Machine Learning Module

December 2025

1 Introduction

Linear boundaries are insufficient for real-world manifolds. In this assignment, you will implement the **RBF (Gaussian) Kernel** to classify concentric and interlocking datasets. You will also learn to perform a **Grid Search** to find the optimal balance between model complexity (γ) and error tolerance (C).

2 The RBF Kernel Logic

Unlike the dot product, the RBF kernel measures similarity based on distance:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

To access this, you can write "rbf" in the SVM function.

3 Task 1: The 2D Visual Experiment

Using the two_moons.csv and the concentric_circles.csv dataset:

1. Fix $C = 10$ and train your SVM with $\gamma = 0.01$, $\gamma = 1$, and $\gamma = 50$.
2. Plot the decision boundaries. Observe how high γ creates "islands" of influence around individual points.

4 Task 2: 3D Hyperparameter Grid Search

To find the best model, we must sweep across two dimensions.

1. Create a grid of $C \in \{0.1, 1, 10, 100\}$ and $\gamma \in \{0.01, 0.1, 1, 10\}$.
2. For each pair (C, γ), calculate the **Validation Error** $(1 - \text{Accuracy})$ on the validation set.
3. **3D Plotting:** Use `plotly.graph_objects.Surface` to plot:
   - X-axis: $\log(C)$
   - Y-axis: $\log(\gamma)$
   - Z-axis: Validation Error