

Software Engineering and Data Science, SEIS 763: Machine Learning

Assignment #8 (100 points)

Due Date: November 15th

The dataset on the Canvas (ML_HW_Data_CellDNA.csv) contains various **numeric** measurements (i.e. size, center, etc) from thousands of bacterium under microscope. The **non-zero** values in the **last column** are the target responses that indicate the bacterium (rows) that are interesting enough for further study. The 0s in the last column indicate the bacterium (rows) are NOT interesting candidates for further study. Convert this target dependent variable to **binary values** of either 0s or 1s for your **two-class** classification.

Write a program using either Python, MatLab, or any programming language of your choice to perform the **two-class** classification analysis on this dataset using the Support Vector Machine method with the “**RBF**” kernel. You do **NOT** need to split data into training and testing sets.

Answer the following questions:

1. Experiment your SVM RBF model with different “**box constraints**” and “**kernel scales / gamma**”.
2. What is the accuracy, Precision, Recall, and F-score for each class prediction under each of your above experiments?
3. Is there any trend that you observed in your experiments?
4. **Optional:** plotting a graph with clear legends and tick labels to illustrate the trend will be very helpful.
5. Create an ROC curve plot for ****EACH**** class in **Just ONE** of your experiments.

Submission Guideline:

1. Please include the WORD document to include your answers (and clearly readable figures/screenshots) to the above questions. Please include **your name** on the top of your WORD document.
2. Please print your program (matlab or python) as **PDF** and include the **PDF** in your submission. Please name your program as “a8.m/.mlx/.py/.inpyb”, depending on the programming language / environment you used.
3. Please also include your program in the formats like .m/.mlx/.py/.inpyb in your submission.
4. Prepare EVERYTHING mentioned in the guideline and submit them on **Canvas** no later than the due date.
5. Please carefully follow the submission guideline. Otherwise, the instructor may not be able to grade your assignment.