PRAGYA VERMA – (101276204)

SEIS 763: 01 Machine Learning

Assignment #5

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

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 (last column) to **binary values** of either 0s or 1s for your **two-class**

classification.

**1. Use \*\*logistic + lasso regression\*\* with \*\*10-fold cross-validation\*\* to**

**identify useful predictors.**

**Ans:** Refer the code

**2. (This question is optional for Python Program) Plot a lasso plot with readable**

**tick labels on the X and Y coordinates in your plot for easy visualization and**

**verification.**

**Ans**: Done in Python

**3. Which top \*\*THREE (3)\*\* remaining predictors (with non-zero theta values) are**

**you going to select to explain why a bacteria is an “interesting” candidates for**

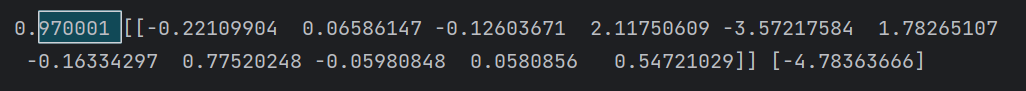
**further study?**

**Ans:** 3, 5 and 7 are the top three predictors (with non-zero theta values) I am going to select for the reason given in the question.

**4. What is the lambda (l) value in Matlab or the *C* value in Python you choose in**

**order to select the top 3 predictors you identified in the last question?**

**Ans:** Chose 0.970001 as the lambda (l) value / C value to choose the top three predictors.

****

**5 - What are the q values for the 3 selected predictors at the lambda (l) value in**

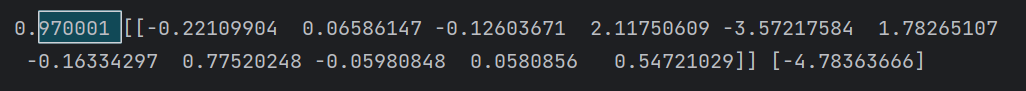
**Matlab or the *C* value in Python you identified in the last question?**

**Ans:** Theta values for the three selected predictors are:

3 - 2.11750609

5 - 1.78265107

7 - 0.77520248

****