Graduate Program in Software
SEIS 763: Machine Learning
Assignment #5 (100 points)
Due Date: July 9th

The dataset on the Blackboard contains various measurements (i.e. size, center, etc) from thousands of bacterium under microscope. The last column with non-zero values indicate the bacterium are interesting enough for further study. Otherwise (i.e. last column with zero values), those bacterium are not interesting candidates for further study. Convert this dependent variable to binary values. Standardize predictors first using Z score.

Write a MatLab (or a programming language of your choice) program to perform an analysis on this dataset using the Support Vector Machine method.

Answer the following questions:

1. How many support vectors did you find?
2. List top 3 records that have the smallest ****absolute**** values from $\boldsymbol{w}^T \bullet \boldsymbol{X} + \boldsymbol{b}$ calculation.
3. What are the "$\boldsymbol{w}^T \bullet \boldsymbol{X} + \boldsymbol{b}$" values for the following records: 131, 165, 892, 1057? Anything special about those values of these few records?

Please follow the instructions below to submit your assignment.

1. If you use Matlab, please name your MatLab program as **"a5.m"**. (or your programming language extension). E-mail the program file(s) to the instructor at clai@stthomas.edu before the class on the due date.
2. Have your program reads in your data from the "U:\tmp" folder.
3. Please submit a hardcopy on the due date to the instructor. Your hardcopy should include (1) your programming solutions (with excellent comments) ****AND**** (2) **a clear and readable screenshot** of your answer for ****EACH**** question. Please also staple all pages of your submission together!!! Instructor is not responsible for missing pages if your submission is not stapled together.

Note: Assignment will be collected right before the class. ***Don't forget to include*** your name on the top of your assignment.