Course: INT 93S
Quarter: Summer 2017
Assigned: 18 July 2017
Due: 9:00 20 July 2017

# Lab 6: Supervised Learning 2

**Collaboration Guidelines: Code and written analysis can be discussed with your lab partner, but must be written separately, (unless you are working on the same code on the same computer together at the same time). Your lab partner is the only student who can look at your code. You can discuss the lab report but it must be written separately.**

*1 point will be taken off for not including a readme and/or improper lab format.*

Turn-in:

1. Create a readme file titled "lastname_firstname_readme.txt" which includes:
    a. First name and last name of you and your partner
    b. How much of the lab you finished in class
    c. Any references you used
    d. Anything you would like the professor and TA to know.
2. Put your code and report in a folder named "lastname_firstname_lab6"
3. Zip the folder and upload it to Gauchospace

Report Format: *(1pt)*

1. In a document file (word, google docs, etc), write up an analysis for the lab.
2. Include your name, the date and the title of the lab.
3. Include the Analysis from part 1 and 2, as well as the graphs.
4. Export Lab report as a pdf saved as "lastname_firstname_lab6_report.pdf"
5. Follow turn-in instructions, including the readme and proper folder/file names

Part 1:

1. Import the dataset (breast-cancer-wisconsin.csv):
2. Reduce the dimensions of the dataset to 2D using PCA and t-SNE and plot the results in a 2X1 grid of graphs with titled subplots.
3. Analyze the results of the two dimension reduction methods using k-Nearest Neighbors and SVM.
    a. Calculate the accuracy of each algorithm for your report.
    b. Graph as a 2X2 grid of graphs with titled subplots.
4. Analysis:
    a. Looking at the chart from step 2, dimension reduction algorithm works best? Why?
    b. What was the accuracy of each algorithm?
    c. Which classifier with which form of dimension reduction worked the best? Why?

Part 2:

*Repeat part 1's analysis on a multidimensional aspect of your data for your research project, your lab partner's dataset, or a dataset of your choice if you are in Group 6.*

1. Import your dataset
2. Reduce the dimensions of the dataset to 2D using PCA and t-SNE and plot the results in a 2X1 grid of graphs with titled subplots.
3. Analyze the results of the two dimension reduction methods using using 2+ supervised machine learning algorithms.
    a. Calculate the accuracy of each algorithm for your report.
    b. Graph as a 2X2 grid of graphs with titled subplots.
4. Analysis:
    a. Looking at the chart from step 2, dimension reduction algorithm works best? Why?
    b. What was the accuracy of each algorithm?
    c. Which classifier with which form of dimension reduction worked the best? Why?