

Project 2 : Somm Body To Love

Due 3/20/2019 @ 12pm (noon)

TOPICS:

- ↪ Python/IPython
- ↪ Logistic Regression
- ↪ Support Vector Machines
- ↪ Decision Trees
- ↪ L2 Regularization
- ↪ Slack Variables
- ↪ Scikit-Learn, Matplotlib



BACKGROUND:

During class, we discussed how to analyze data to make a prediction based on a set of features using a logistic regression, support vector machines, decision trees. A major issue associated with supervised machine learning algorithms is that the models developed by the training set is overfitted and not generalizable for an independent test set. L2 regularization and slack variables can reduce overfitting so that the model can more generally classify data correctly to a test set.

DIRECTIONS:

In this project, you will use a logistic regression, support vector machine, and decision tree algorithms to analyze the entire set of features from a wine dataset (<https://archive.ics.uci.edu/ml/machine-learning-databases/wine/wine.data>) to predict to predict the wine's alcohol type.

You will use library functions to implement the logistic regression, support vector machines, and decision tree algorithms on a training set and apply the model to a separate test set. You will calculate the training error and using L2 regularization or slack variables to obtain an optimal set of predictions.

You will calculate and graph (within the IPython Notebook) the test error vs. lambda or C, depending on whether you are using logistic regression or support vector machines, respectively. Note that there is no template for this in the Chapter 3 Jupyter notebook from your textbook.

IMPLEMENTATION NOTES:

Any program that does not execute completely without errors will not be graded.

COMMENTS AND STYLE:

Although there will be no formal policy on commenting and style, the reader should be able to easily follow the main purpose of the code. Each set of code that does something significant must be commented. The variable names should be easily recognizable and acronyms should be avoided if possible.

Do not be surprised if help is not forthcoming if your code is poorly commented and/or difficult to follow. You have been warned.

PROJECT SUBMISSION:

You will turn in the modified IPython notebook with a different set of features.

The programs and graphs should be in a single directory named "SommBody". The contents of the directories must be archived in a tarball that is gzipped called Proj2.tar.gz.

Place the gzipped tarball in your Drop Box on Sakai before it is due.

PLEDGED WORK POLICY:

Assignments in Computer Science courses may be specified as "pledged work" assignments by the professor of the course. When an assignment is specified as "pledged work" the only aid that the student may seek is from either the course professor or TAs (including CS Center tutors) that the professor has explicitly specified. On "pledged work" assignments the student may not use the services of a tutor.

For this project, you and your partner will develop code together into shared repositories, so you will see and share work. In addition, you may discuss only **basic programming language syntax and general computer science concepts** with everyone else. Any other communications of the project (e.g., giving your code to someone else or seeing someone else's code) are strictly prohibited except with the professor and TAs of the course. Your code and your implementation of the project must be the product of your own work and that of your partner.