

**CIS 4526 Foundations of Machine Learning**

**Logistic Regression  
&  
Support Vector Machines**

Khai Nguyen

# CIS 4526 Foundations of Machine Learning

## 1. Design

### a. Core functions

- i. **train\_classifier**(train\_x, train\_y, learn\_rate, loss, lambda\_val=None, regularizer=None):
  - initialize delta\_weight vector
  - calculate current\_loss
  - loop through length of delta\_weight
    - calculate temp\_loss (h) for each index of delta\_weight
    - calculate delta\_weight[index] = (temp\_loss - current\_loss)/h
  - return weight vector
- ii. **test\_classifier**(w, test\_x)
  - take dot product of weight\_vector and each row of testX
  - return predictY
- iii. **normalize**(trainX, testX)
  - normalize data based on “mean” and “stdev”
- iv. **compute\_accuracy**(test\_y, pred\_y)

### b. Logistic Regression

- i. **logistic\_loss**(train\_y, pred\_y):
  - $\log\_loss = \log(1 + \exp(-yy'))$
- ii. **weight\_vector\_logistic** = train\_classifier(trainX, trainY, 0.001, logistic\_loss)

### c. SVM

- i. **hinge\_loss**(train\_y, pred\_y)
  - $\text{hinge\_lossVector} = \max(0, 1 - yy')$
- ii. **l2\_reg**(w)
- iii. **weight\_vector\_SVM** = train\_classifier(trainX, trainY, 0.1, hinge\_loss, 0.001, l2\_reg)

### d. main()

- 5-fold splitting
- Normalize
- Compute weight vector for Logistic Regression & SVM
- Compute pred\_y for Logistic Regression & SVM
- Compute accuracy

# CIS 4526 Foundations of Machine Learning

## 2. Results

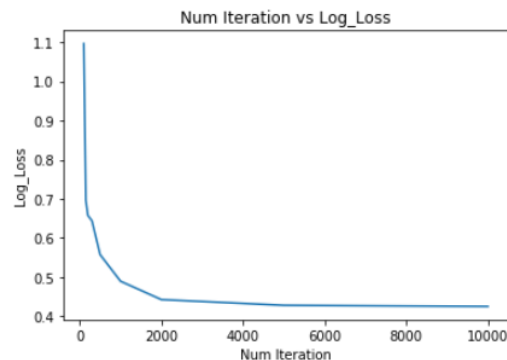
Below is the table showing the result (with 5-fold validation)

0 th iteration
Accuracy is of Logistic Regression
0.6092592592592593
Accuracy is of SVM
0.6129629629629629
1 th iteration
Accuracy is of Logistic Regression
0.5962962962962963
Accuracy is of SVM
0.6018518518518519
2 th iteration
Accuracy is of Logistic Regression
0.7240740740740741
Accuracy is of SVM
0.7240740740740741
3 th iteration
Accuracy is of Logistic Regression
0.4555555555555555
Accuracy is of SVM
0.4555555555555555
4 th iteration
Accuracy is of Logistic Regression
0.6370370370370371
Accuracy is of SVM
0.6425925925925926

## 3. Discussion:

Looking at the results for both algorithms, most sup-samples in the 5-fold cross validation have the same accuracy. Therefore, Logistic Regression and SVM both have similar performance as classifiers.

For one experiment with logistic regression, a plot of the loss vs. the number of iterations is displayed below.



## CIS 4526 Foundations of Machine Learning

In addition, by plotting the curve between **Accuracy** vs. **Number of Iterations**, we can see that after around 800 iterations, the value of accuracy stabilizes. Therefore, we can choose any number of iterations larger than 800, but for the purpose of not creating extra computation time but and ensuring the loss function is also stabilized (around **2000**), we can pick **2000** as the number of iterations.

