

# CS 273A Machine Learning Homework 1

Chen Li

October 8, 2015

## 1 Problem1 Perceptron learning

### 1.1 1(a)

1. I obtain two cluster center from a 9-dimensional spherical Gaussian distribution with  $\mu$  equal to 0,  $\sigma$  equal to 1.
2. Let  $\sigma^* = \alpha \times \sigma$  and  $\mu$  equal to the two cluster centers obtained from step 1, I obtain 2 classes of 9-dimensional input vectors.

### 1.2 1(b)

1. First obtain the mean  $m1$  and  $m2$  of the two clusters, then calculate the total within-class covariance matrix  $S_w$ , finally get the direction  $w \propto S_w^{-1}(m2 - m1)$ . Then set threshold to be  $w \cdot \frac{(m1+m2)}{2}$ , i.e., let the line pass the middle point of  $m1$  and  $m2$ .
2. Then I vary  $\alpha$  from 0.1 to 0.9,  $N$  from 100 to 1000. Each combination I give 100 tests to obtain the mean and standard deviation of accuracy.

### 1.3 1(c)

1. Firstly I use Linear regression on the two data sets to generate the initial  $w_0$ , then I use stochastic gradient decent in certain numbers of iterations to obtain the  $w$ .
2. Set  $\alpha = 0.4$  and  $N = 500$ , I use the similar approach in 1(b) on *iteration* and  $N_{test}$  to obtain the mean and standard deviation of accuracy.

## 2 Problem2 Toy OCR

### 2.1 2(a)

Set two 9-dimensional zero vectors and randomly choose  $k$  (from 2 to 9) position to fill with 1 as the two cluster center. Other part use the same approach as problem 1.

## 3 Implementation & Result Analysis

In *hw1.py*, the *genData()* and *genOCRData()* function respectively implements the approaches in 1(a) and 2(a), sample datas locate in the folder *sample\_data*. The *Fisher()* function implements the Fisher's Linear Discrimination, and the *FisherTest()* function test the *Fisher()* function in different combinations of  $\alpha$  and  $N$ . Similarly, the *Perceptron()* function implements the Perceptron Learning Algorithm, and the *PLATest()* function test the *Perceptron()* function in different combinations of *iteration* and  $N_{test}$ . The result of the test can be found in the *test\_results* folder.

Overall, the result of PLA is better than Fisher. For fisher, when  $\alpha$  and  $N$  increase, there error and error bar slightly increase. For PLA, the iteration time and test size don't infect the result much. Plots can be find in In *plot\_results* folder.