CS440 : Introduction to Artificial Intelligence	**Spring 2014**
Problem Set 2 - Part 2	
Dan McQuillan	*Handed In: February 15, 2014*

1. **Solution to problem 2**
The weight vector values were largely different as shown below:

old	**new**
$\theta = -54.0$	$\theta = 513.0$
$w_1 \to -752.0$	$w_1 \to -1719.0$
$w_2 \to 4771.0$	$w_2 \to 2216.0$
$w_3 \to 17714.0$	$w_3 \to 1357.0$
$w_4 \to 762.0$	$w_4 \to 1506.0$
$w_5 \to 6.0$	$w_5 \to 435.0$
$w_6 \to 676.0$	$w_6 \to 1152.0$
$w_7 \to 3060.0$	$w_7 \to -300.0$
$w_8 \to -2004.0$	$w_8 \to -1928.0$
$w_9 \to 5459.0$	$w_9 \to 1462.0$
$w_{10} \to 9591.2999999989$	$w_{10} \to 991.9000000000113$
$w_{11} \to 3832.0$	$w_{11} \to -775.0$
$w_{12} \to 14963.0$	$w_{12} \to 6726.0$
$w_{13} \to 20912.0$	$w_{13} \to 11721.0$

The confusion matrix also was also more error prone when testing data set 1. It resulted in the following confusion matrix:

53	1
9	54

The application to data set 3 produced largely different conclusions.

However, the most important feature's weight value stayed relatively the same which shows that it has more influence on the Δw on each iteration of the inputs. This is as expected since because that feature is the most important.

2. **Solution to problem 3**
   (a) Find and report the list MAX for Dataset 2 $(1.0, 4.0, 170.0, 407.0, 1.0, 2.0, 194.0, 1.0, 4.0, 3.0, 3.0, 7.0)$
   (b) Repeat Problem 1

   Training Dataset 1x

The training took 14 epochs.
$Threshhold : 5.0$
$\gamma \to 0.0320963696907362$
$w_1 \to -1.2368421052631573$
$w_2 \to 3.0$
$w_3 \to 5.5$
$w_4 \to -0.2899999999999997$
$w_5 \to -0.13829787234042568$
$w_6 \to 0.0$
$w_7 \to 0.5$
$w_8 \to -13.317204301075275$
$w_9 \to 3.0$
$w_{10} \to 5.645161290322578$
$w_{11} \to 2.9999999999999987$
$w_{12} \to 5.0$
$w_{13} \to 5.7142857142857135$

Test Dataset 1x

Confusion Matrix:

54	0
0	63

Total loss: 0.0

Test Dataset 2x

False Positives:
Index: 7
Confusion matrix:

13	0
1	19

Total loss: 0.010623925503195387

Application Dataset 3x

$1 \to 0.0$
$2 \to 1.0$
$3 \to 0.0$
$4 \to 1.0$
$5 \to 1.0$
$6 \to 0.0$

$7 \rightarrow 0.0$
$8 \rightarrow 0.0$
$9 \rightarrow 0.0$
$10 \rightarrow 1.0$
$11 \rightarrow 1.0$
$12 \rightarrow 0.0$
$13 \rightarrow 1.0$
$14 \rightarrow 0.0$
$15 \rightarrow 0.0$
$16 \rightarrow 0.0$
$17 \rightarrow 1.0$
$18 \rightarrow 0.0$
$19 \rightarrow 0.0$
$20 \rightarrow 0.0$
$21 \rightarrow 0.0$

(c) **Explanation**
The results above show that the weights were quite significantly smaller and the number of epoch were also quite a bit less. This is a result of the perceptron having to oscillate because of the larger difference between input vectors. Instead it is more likely to increase with a smaller Δw and have a lesser chance to have to use a negative Δw to correct a movement when it moves past the margin of a "good" perceptron.

(d) **Test new perceptron on DataSet1**
When testing data set 1 it resulted in a complete failure of classifying the input vectors from dataset one. This is a result of normalizing the input vectors before training which therefore resulted in a normalization of the weight vector. This caused the perceptron to misclassify one of the classifications since it resulted in being below all the points and caused them all to be classified as positive.

(e) **Is is possible to recover the problem 1 perceptron**
It is not possible to recover the perceptron from problem 1 from this weight vector since the resultant weight vectors are dependent on the values chosen to be the initial training weights. This is because the values for the applied Δw's will be different for each set of weights that are applied. If the weights is higher for example then the training may be more likely to jump past the set of "good" perceptrons.

3. **Solution to problem 4**

(a) Train on DataSet4, test on DataSet1
Training Results:
The training took 2443 epochs.
Threshhold: 413.0
$\gamma \rightarrow 0.01087449395327017$

$w_1 \to -39.0$
$w_2 \to 3052.0$
$w_3 \to 3274.0$
$w_4 \to 533.0$
$w_5 \to 550.0$
$w_6 \to 2354.0$
$w_7 \to -3552.0$
$w_8 \to -2025.0$
$w_9 \to 2148.0$
$w_{10} \to 2022.2999999999358$
$w_{11} \to -1500.0$
$w_{12} \to 6491.0$
$w_{13} \to 14862.0$

Testing Results
False Positives:
Index: 5
Inputs: $(67.0, 0.0, 3.0, 115.0, 564.0, 0.0, 2.0, 160.0, 0.0, 1.6, 2.0, 0.0, 7.0)$
Index: 6
Inputs: $(65.0, 0.0, 3.0, 140.0, 417.0, 1.0, 2.0, 157.0, 0.0, 0.8, 1.0, 1.0, 3.0)$
Index: 8
Inputs: $(65.0, 0.0, 3.0, 160.0, 360.0, 0.0, 2.0, 151.0, 0.0, 0.8, 1.0, 0.0, 3.0)$
Index: 20
Inputs: $(64.0, 0.0, 3.0, 140.0, 313.0, 0.0, 0.0, 133.0, 0.0, 0.2, 1.0, 0.0, 7.0)$
Index: 33
Inputs: $(64.0, 0.0, 4.0, 180.0, 325.0, 0.0, 0.0, 154.0, 1.0, 0.0, 1.0, 0.0, 3.0)$
Index: 53
Inputs: $(74.0, 0.0, 2.0, 120.0, 269.0, 0.0, 2.0, 121.0, 1.0, 0.2, 1.0, 1.0, 3.0)$
Index: 77
Inputs: $(66.0, 0.0, 1.0, 150.0, 226.0, 0.0, 0.0, 114.0, 0.0, 2.6, 3.0, 0.0, 3.0)$
Index: 114
Inputs: $(60.0, 0.0, 3.0, 120.0, 178.0, 1.0, 0.0, 96.0, 0.0, 0.0, 1.0, 0.0, 3.0)$

False Negatives:
Index: 34
Inputs: $(53.0, 1.0, 4.0, 140.0, 203.0, 1.0, 2.0, 155.0, 1.0, 3.1, 3.0, 0.0, 7.0)$

Confusion Matrix:

53	1
8	55

Total loss: 21.525012758300132

(b) **Train on DataSet5, test on DataSet1**
Training Results:
The training took 1190 epochs.
Threshhold: 272.0
$\gamma \to 0.04518422260458557$

$w_1 \to 177.0$
$w_2 \to 2699.0$
$w_3 \to 819.0$
$w_4 \to 631.0$
$w_5 \to 497.0$
$w_6 \to 1189.0$
$w_7 \to -952.0$
$w_8 \to -1948.0$
$w_9 \to 2047.0$
$w_{10} \to 1521.3999999999726$
$w_{11} \to -1380.0$
$w_{12} \to 5972.0$
$w_{13} \to 10336.0$

Testing Results
False Positives:
Index: 5
Inputs: $(67.0, 0.0, 3.0, 115.0, 564.0, 0.0, 2.0, 160.0, 0.0, 1.6, 2.0, 0.0, 7.0)$
Index: 6
Inputs: $(65.0, 0.0, 3.0, 140.0, 417.0, 1.0, 2.0, 157.0, 0.0, 0.8, 1.0, 1.0, 3.0)$
Index: 8
Inputs: $(65.0, 0.0, 3.0, 160.0, 360.0, 0.0, 2.0, 151.0, 0.0, 0.8, 1.0, 0.0, 3.0)$
Index: 20
Inputs: $(64.0, 0.0, 3.0, 140.0, 313.0, 0.0, 0.0, 133.0, 0.0, 0.2, 1.0, 0.0, 7.0)$
Index: 31
Inputs: $(69.0, 1.0, 1.0, 160.0, 234.0, 1.0, 2.0, 131.0, 0.0, 0.1, 2.0, 1.0, 3.0)$
Index: 33
Inputs: $(64.0, 0.0, 4.0, 180.0, 325.0, 0.0, 0.0, 154.0, 1.0, 0.0, 1.0, 0.0, 3.0)$
Index: 53
Inputs: $(74.0, 0.0, 2.0, 120.0, 269.0, 0.0, 2.0, 121.0, 1.0, 0.2, 1.0, 1.0, 3.0)$
Index: 77
Inputs: $(66.0, 0.0, 1.0, 150.0, 226.0, 0.0, 0.0, 114.0, 0.0, 2.6, 3.0, 0.0, 3.0)$
Index: 91
Inputs: $(76.0, 0.0, 3.0, 140.0, 197.0, 0.0, 1.0, 116.0, 0.0, 1.1, 2.0, 0.0, 3.0)$
Index: 114
Inputs: $(60.0, 0.0, 3.0, 120.0, 178.0, 1.0, 0.0, 96.0, 0.0, 0.0, 1.0, 0.0, 3.0)$

False Negatives:
Index: 34
Inputs: $(53.0, 1.0, 4.0, 140.0, 203.0, 1.0, 2.0, 155.0, 1.0, 3.1, 3.0, 0.0, 7.0)$
Index: 69
Inputs: $(60.0, 1.0, 4.0, 117.0, 230.0, 1.0, 0.0, 160.0, 1.0, 1.4, 1.0, 2.0, 7.0)$
Index: 95
Inputs: $(39.0, 1.0, 4.0, 118.0, 219.0, 0.0, 0.0, 140.0, 0.0, 1.2, 2.0, 0.0, 7.0)$
Confusion Matrix:

51	3
10	53

Total loss: 32.68215493625958

(c) Train on DataSet6, test on DataSet1
Training Results:
The training took 2593 epochs.
Threshhold: 858.0
$\gamma \to 0.022928058158089223$

$w_1 \to -1853.0$
$w_2 \to 3172.0$
$w_3 \to 3650.0$
$w_4 \to 914.0$
$w_5 \to 588.0$
$w_6 \to 753.0$
$w_7 \to -3254.0$
$w_8 \to -1900.0$
$w_9 \to 1060.0$
$w_{10} \to 1868.8000000000209$
$w_{11} \to -580.0$
$w_{12} \to 8319.0$
$w_{13} \to 17636.0$

Testing Results
False Positives:
Index: 5
Inputs: $(67.0, 0.0, 3.0, 115.0, 564.0, 0.0, 2.0, 160.0, 0.0, 1.6, 2.0, 0.0, 7.0)$
Index: 6
Inputs: $(65.0, 0.0, 3.0, 140.0, 417.0, 1.0, 2.0, 157.0, 0.0, 0.8, 1.0, 1.0, 3.0)$
Index: 8
Inputs: $(65.0, 0.0, 3.0, 160.0, 360.0, 0.0, 2.0, 151.0, 0.0, 0.8, 1.0, 0.0, 3.0)$
Index: 20
Inputs: $(64.0, 0.0, 3.0, 140.0, 313.0, 0.0, 0.0, 133.0, 0.0, 0.2, 1.0, 0.0, 7.0)$
Index: 33

Inputs: $(64.0, 0.0, 4.0, 180.0, 325.0, 0.0, 0.0, 154.0, 1.0, 0.0, 1.0, 0.0, 3.0)$

False Negatives:
Index: 34
Inputs: $(53.0, 1.0, 4.0, 140.0, 203.0, 1.0, 2.0, 155.0, 1.0, 3.1, 3.0, 0.0, 7.0)$
Index: 69
Inputs: $(60.0, 1.0, 4.0, 117.0, 230.0, 1.0, 0.0, 160.0, 1.0, 1.4, 1.0, 2.0, 7.0)$
Index: 108
Inputs: $(54.0, 1.0, 4.0, 110.0, 206.0, 0.0, 2.0, 108.0, 1.0, 0.0, 2.0, 1.0, 3.0)$
Confusion Matrix:

51	3
5	58

Total loss: 13.610748698886589

(d) Which of the four yields the best performance?
The test training from the permutation of the second dataset had the best performance since we are looking at minimum loss as a determining factor in the success of the perceptron. The results will not always be identical since for the most part the permutations are generated via random numbers and since they are seeded via the time for the Collections.shuffle method it will produce different results each run of the training.

(e) Test averaged weight vector on dataset 1.
Training on dataset2:
Threshhold: 513.0
$\gamma \rightarrow 2.7031724654225244 \times 10^{-4}$
$w_1 \rightarrow -1719.0$
$w_2 \rightarrow 2216.0$
$w_3 \rightarrow 1357.0$
$w_4 \rightarrow 1506.0$
$w_5 \rightarrow 435.0$
$w_6 \rightarrow 1152.0$
$w_7 \rightarrow -300.0$
$w_8 \rightarrow -1928.0$
$w_9 \rightarrow 1462.0$
$w_{10} \rightarrow 991.9000000000113$
$w_{11} \rightarrow -775.0$
$w_{12} \rightarrow 6726.0$
$w_{13} \rightarrow 11721.0$

Averaged weight vector:

$threshold \to 0.030422435180373\,23$
$w_1 \to -0.04957994461909904$
$w_2 \to 0.17236419772286374$
$w_3 \to 0.12970485078786842$
$w_4 \to 0.05714024689694883$
$w_5 \to 0.03208977822798198$
$w_6 \to 0.08540346718962719$
$w_7 \to -0.11290404208868643$
$w_8 \to -0.12295889110966987$
$w_9 \to 0.1083073557681354$
$w_{10} \to 0.09787094960776252$
$w_{11} \to -0.06843234083135682$
$w_{12} \to 0.42531877515611183$
$w_{13} \to 0.827552512265163$

Results:
False Positives:
Index: 5
Inputs: $(67.0, 0.0, 3.0, 115.0, 564.0, 0.0, 2.0, 160.0, 0.0, 1.6, 2.0, 0.0, 7.0)$
Index: 6
Inputs: $(65.0, 0.0, 3.0, 140.0, 417.0, 1.0, 2.0, 157.0, 0.0, 0.8, 1.0, 1.0, 3.0)$
Index: 8
Inputs: $(65.0, 0.0, 3.0, 160.0, 360.0, 0.0, 2.0, 151.0, 0.0, 0.8, 1.0, 0.0, 3.0)$
Index: 20
Inputs: $(64.0, 0.0, 3.0, 140.0, 313.0, 0.0, 0.0, 133.0, 0.0, 0.2, 1.0, 0.0, 7.0)$
Index: 31
Inputs: $(69.0, 1.0, 1.0, 160.0, 234.0, 1.0, 2.0, 131.0, 0.0, 0.1, 2.0, 1.0, 3.0)$
Index: 33
Inputs: $(64.0, 0.0, 4.0, 180.0, 325.0, 0.0, 0.0, 154.0, 1.0, 0.0, 1.0, 0.0, 3.0)$
Index: 77
Inputs: $(66.0, 0.0, 1.0, 150.0, 226.0, 0.0, 0.0, 114.0, 0.0, 2.6, 3.0, 0.0, 3.0)$
Index: 114
Inputs: $(60.0, 0.0, 3.0, 120.0, 178.0, 1.0, 0.0, 96.0, 0.0, 0.0, 1.0, 0.0, 3.0)$

False Negatives:
Index: 34
Inputs: $(53.0, 1.0, 4.0, 140.0, 203.0, 1.0, 2.0, 155.0, 1.0, 3.1, 3.0, 0.0, 7.0)$
Index: 69
Inputs: $(60.0, 1.0, 4.0, 117.0, 230.0, 1.0, 0.0, 160.0, 1.0, 1.4, 1.0, 2.0, 7.0)$

Confusion Matrix:

52	2
8	55

Total loss: 21.093763810153984

The results were successful in classifying data set 1 with minimal error. Also, the total loss ended up being near the average of the losses of datasets 2,4,5, and 6.