Pattern Informatics Report

Le Trung Kien
03-120291, 3rd Year
Department of Mechano-Informatics
The University of Tokyo

February 17, 2013

1 Widrow-Hoff Algorithm & Pseudo-Inverse Matrix

Widrow-Hoff Algorithm based on gradient descent method to update model's weight as following:

$$W_{new} = W_{old} - \alpha W_{old}(XW_{old} - T)$$

In which, α is learning rate, W is weight, X is training data, T is training label. The training will stop when error reaches a threshold or number of iterations exeeds limit. In my implementation, learning rate is 0.1, threshold is 0.001 and max number of iterations is 100.
On the other hand, optimal value can be directly calculated by using pseudo-inverse matrix:

$$W = (X^T X)^{-1} X^T T$$

Results after using the two above algorithms are shown in Figure 1.

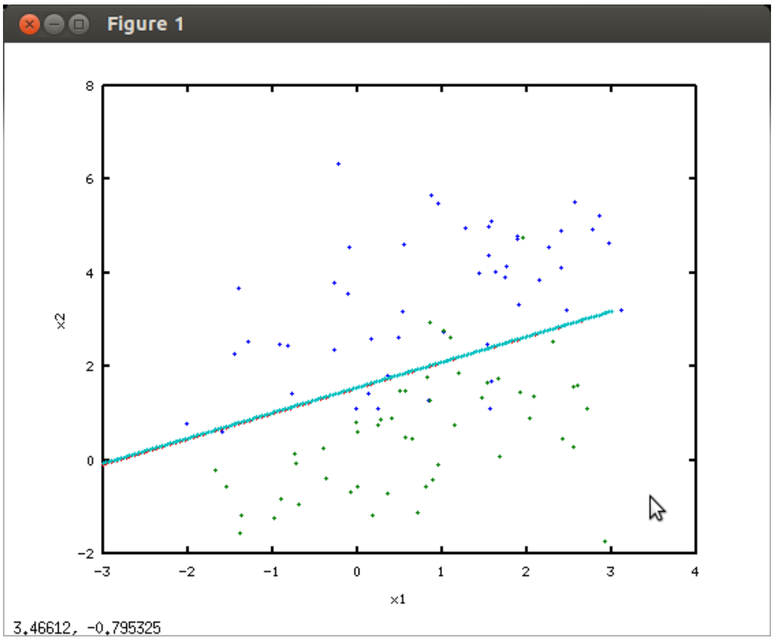

Figure 1: Classification by Widrow-Hoff algorithm and pseudo-inverse matrix.

According to Figure 1, the two weights trained by the two algorithms are almost the same, which is not obvious fact. In this case, the two classes are linearly separable, so the two algorithms should lead to very simal final weight. Still, they are slightly different because the weight calculated by pseudo-inverse matrix is deterministic while the one trained by Willow-Hoff depends on numerous factors such as intial values, error's threshold and number of iterations. Moreover, Willow-Hoff algorithm is always a valid method, but when pseudo-inverse matrix of X does not exist, we cannot use it to find the optimal weight for classification.

An example of running the program (Willow-Hoff and Pseudo-Inverse in order):

```
@:~/Pattern$ ./2 1
0.642855
0.223755
-0.412262
Precision’s rate: 0.875
@:~/Pattern$ ./2 2
0.644933
0.223556
-0.412747
Precision’s rate: 0.875
```

2 Leave-one-out Method

3 Mahalanobis Distance