Weicong Feng

Professor Chia-Ling Tsai

CSC74020 Machine Learning

10/13/2021

Assignment2

**Report**

# Solution

Basically, the pocket algorithm is a variant of PLA. The main difference is that pocket evaluate the loss using whole training data set after each updating weights and keep the best weights, while PLA doesn’t. Hence, mathematically, they have same formular:

Text

Description automatically generated

The weights are updated:

The weights will turn to the correct direction when the weights are updated, as shown below:

Chart, radar chart

Description automatically generated

# Training and validation

## Experiment

In this homework, the PLA is created as a parent class, and the pocket class inherits from PLA. Hence, two algorithms are compared. In each algorithm, 3 weights initialization approaches are used and compared, the first entry of training data, randomly, and linear regression with pseudo-inverse.

In terms of the setup of data set, the whole data set of 569 entries were divided to 5 parts, each part including 100 entries. Each environment, different algorithms and weights initialization ways, will be run 5 times, where each time use different data set. For instance, 0-99 entries in data set will be used as validation data set (Eout) while 100-499 entries in data set will be used for training (Ein) in the first time; 100-199 entries in data set (Eout) will be used as validation data set while 0-99 and 200-499 entries in data set will be used for training (Ein) in the second time; and so on. For each running, 50, 000 epochs or error rate less than 0.01.

In each running, each 500 epochs, the Eout and Ein are recorded. After the experiment, the average of Eout and Ein are calculated respectively.

## Result

Graphical user interface, application

Description automatically generated

Above charts are the result of this experiment.

## Discussion

From the performance charts, we can conclude that the loss of PLA fluctuates seriously than the loss of pocket algorithm. Among various initialization approaches, there isn’t big differences. Theoretically, the linear regression initialization should cause reaching the best performance and low loss very quickly. However, even other initializations reach the best performance in a few epochs, so it is difficult to show in the charts. The common is that the valuation loss always higher than the loss of the training loss after first epoch. In addition, after dozens of epochs training, the Ein continue going down while Eout going up, it means overfitting. Because the PLA doesn’t store the best weights, its best performance appears in the first epoch, after weights updating, the loss rises quickly.