Perceptron Summary

Bill Roland

March 3, 2018

1 Context

Perceptron performs well when the dataset is linearly seperable.In this experiment, I use a dataset which is non-linearly seperable to test perceptron.

2 Equation Derivation

Perceptron use a hyperplane to seperate dataset into two class.Several samples will be misclasified until the proper hyperplane has been found(only if the dataset is linear seperable).At each epoch, use these misclassified sample to adjust hyperplane i.e. minimize the loss function).

Loss funtion:(M is a set of misclassified samples)

$$L\left(w,b\right)=-\sum_{x_i\in M}y^{(i)}\left(w^Tx_i+b\right)$$

Use batch gradient descent to minimize loss funtion:

$$\frac{\partial L}{w}=-\sum_{x_i\in M}y^{(i)}x_i$$
$$\frac{\partial L}{b}=-\sum_{x_i\in M}y^{(i)}$$

Update parameters:

$$w:=w+\eta y^{(i)}x_i$$
$$b:=b+\eta y^{(i)}$$

3 Experiment

3.1 Introduction

Here is a dataset about breast cancer [1]. Use perceptron to classify this dataset.Select model from 10 of them based on the result of 4-fold cross val-

[1]https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)

idation.As this dataset is nonlinear seperable, stop the training process when misclassified samples are less than 10% of training set.

3.2 Result

The error percentage is as low as aroud 0.9%, and precison can be 95% while recall is 63%.

4 Summary

Perceptron is easy to implement and has good performance on binary classfication task.