An Introduction to the Perceptron Algorithm

Haoji Hu

22 November 2011

1 The Problem

Suppose there are N vectors in a p dimensional vector space. Each of these N vectors belongs to either one of two classes C_1 and C_2. Our task is to find a hyperplane to separate these vectors out based on the classes they belong to.

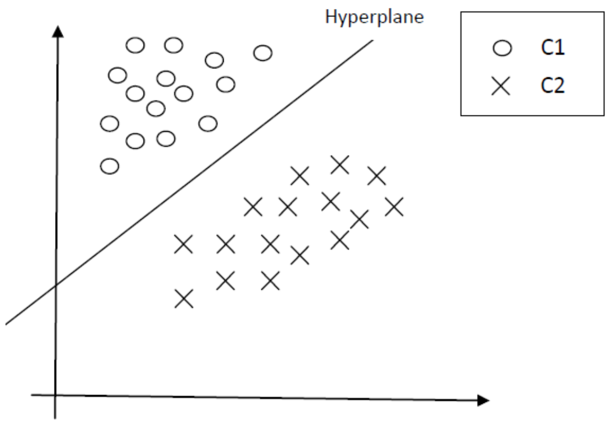

Figure 1: A hyperplane (or a line) which separates two-dimensional vectors into two classes.

2 The Mathematical Description

Here we formulate the problem into a more mathematical form. Suppose x_1, x_2, $\cdots$, x_N are p dimensional vectors. Let's define their extended vectors, $\overrightarrow{x_1}$, $\overrightarrow{x_2}$, $\cdots$, $\overrightarrow{x_N}$, as follows:

$$\text{If } x_i \in C_1,\ \text{then } \overrightarrow{x_i} = \begin{bmatrix} x_i \\ 1 \end{bmatrix};\ \text{if } x_i \in C_2,\ \text{then } \overrightarrow{x_i} = \begin{bmatrix} -x_i \\ -1 \end{bmatrix}$$

All extended vectors are $p+1$ dimensional. The problem has been changed as finding a $p+1$ dimensional vector ω such that for each $i = 1, 2, ..., N$,

$$\omega^T \overrightarrow{x_i} > 0$$

3 The Algorithm

The perceptron algorithm has been proposed by Frank Rosenblatt in 1956. It is simple, but significant. The algorithm has preluded the area of machine learning and pattern recognition. The procedure of the algorithm is shown in Algorithm 1.

Input: $\overrightarrow{x_1}, \overrightarrow{x_2}, \cdots, \overrightarrow{x_N}$
Output: w

$w = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$;
$FLAG = 1$;
while $FLAG$ **do**
    $FLAG = 0$;
    for i=$1{:}N$ **do**
        if $\omega^T \overrightarrow{x_i} \leq 0$ **then**
            $\omega = \omega + \overrightarrow{x_i}$;
            $FLAG = 1$
        end
    end
end
return ω;

Algorithm 1: The Perceptron Algorithm

4 The Proof of the Algorithm's Convergence

Theorem 1 *For N vectors $\overrightarrow{x_1}, \overrightarrow{x_2}, \cdots, \overrightarrow{x_N}$ in the vector space, if there exists an vector ω_{opt} such that $\omega_{opt}^T \overrightarrow{x_i} > 0$ for all $i = \{1, 2, ..., N\}$, then the perceptron algorithm described in Algorithm 1 could finally find a vector ω such that $\omega^T \overrightarrow{x_i} > 0$ for all $i = \{1, 2, ..., N\}$. The convergence does not depend on the initial selection of ω.*

Proof Without loss of generality, we suppose that $\|\omega_{opt}\| = 1$ (Think about why we could make such an assumption).

We further define $\omega(k)$ as the ω at the kth iteration. Then there are two cases:

1. If $\omega(k)^T \overrightarrow{x_i} > 0$ for all $i = \{1, 2, ..., N\}$, the theorem has been proved.

2. Otherwise, there exists at least one $i \in \{1, 2, ..., N\}$ which makes $\omega(k)^T \overrightarrow{x_i} \leq 0$. Then based on the perceptron algorithm,

$$\omega(k+1) = \omega(k) + \overrightarrow{x_i}$$

, which means that

$$\omega(k+1) - a\omega_{opt} = \omega(k+1) - a\omega_{opt} + x_i$$

Taking the module of each side, we obtain

$$\begin{aligned} &\|\omega(k+1) - a\omega_{opt}\|^2 \\ =&\|\omega(k) - a\omega_{opt} + \overrightarrow{x_i}\|^2 \\ =&\|\omega(k) - a\omega_{opt}\|^2 + \|\overrightarrow{x_i}\|^2 + 2\omega(k)^T \overrightarrow{x_i} - 2a\omega_{opt}^T \overrightarrow{x_i} \end{aligned}$$

Here a is a positive number which we are going to discuss later. Because $\omega(k)^T \overrightarrow{x_i} \leq 0$, we can have:

$$\begin{aligned} &\|\omega(k+1) - a\omega_{opt}\|^2 \\ \leq&\|\omega(k) - a\omega_{opt}\|^2 + \|\overrightarrow{x_i}\|^2 - 2a\omega_{opt}^T \overrightarrow{x_i} \end{aligned}$$

Here we define $\beta = \max_{i=1}^N \|\overrightarrow{x_i}\|$, and $\gamma = \min_{i=1}^N (\omega_{opt}^T \overrightarrow{x_i})$ ($\beta > 0$ and $\gamma > 0$, why?). Then it is easy to prove that when $a = \frac{\beta^2+1}{2\gamma}$,

$$\|\omega(k+1) - a\omega_{opt}\|^2 \leq \|\omega(k) - a\omega_{opt}\|^2 - 1 \tag{1}$$

If we take $a = \frac{\beta^2+1}{2\gamma}$, then $\|a\omega_{opt}\| = a = \frac{\beta^2+1}{2\gamma}$. Define the distance between the initial vector $\omega(0)$ and $a\omega_{opt}$ as $D = \|\omega(0) - a\omega_{opt}\|$. Based on Eqn (1), for each iteration, the distance from ω to $a\omega_{opt}$ has been decreased at least 1, so for at most D^2 iterations, the vector ω will converge to $a\omega_{opt}$.

Thus, we have finished the proof.