Lecture 7, Exercise 1

December 13, 2017

Author: Alan Le Goallec

1 Question

The Worst Case of Perceptron
Source: Shalev-Shwartz, Shai, and Shai Ben-David. Understanding machine learning: From theory to algorithms.
For any positive integer m, find a sequence of examples $\{(x_1, y_1), ..., (x_m, y_m)\}$ such that when running the perceptron on this sequence of examples starting from w(0) = 0, it makes at least m updates before converging.

Hint: Try to use training data drawn from $R^m, i.e., x_i \in R^m$.

2 Solution

Let us construct the following data: for every i, we let $x_i = e_i$, the standard basis of i in R^m.
First note that if we pick $w = (y_1, y_2, ..., y_m)^T$, we have for any i, $y_i x_i^T w = y_i e_i^T w = y_i^2 = 1 > 0$.
So this example is indeed linearly separable and the perceptron converges to a w^*.
When the perceptron converges we have $y_i x_i^T w^* > 0$,for every i.
Since y_i is either 1 or -1, we have that for any i, $x_i^T w^* = e_i^T w^* = (w^*)_i \neq 0$.
That is, the solution w^* after running the perceptron algorithm must be nonzero for every coordinate.
Note that when running perceptron, we update w when there exists i such that $y_i x_i^T w(t) < 0$, and w is updated as follows:
$w(t+1) = w(t) + \eta y_i x_i$.
Since $x_i = e_i$, it is immediate that the above update only changes the ith coordinate of w(t). Note that initially, w(0) = 0. Hence, in order to update it to become a vector such that each of its coordinate is nonzero, we need at least m updates. Therefore, the perceptron needs at least m updates to converge.