Intelligence Artificial and the Web Exam

MOSIG-2013

Massih-Reza Amini (Part 1), Ahlame Douzal (Part 2)
Duration: 3 hours, Documents: authorized

Part 1: An analysis of the perceptron algorithm (10 Pts)

The perceptron algorithm is one of the first supervised models proposed by Rosenblatt, 1957 for binary classification. The training step of the algorithm consists in finding the parameters of a linear function defined by

$$\begin{aligned} h_w : \mathbb{R}^d &\rightarrow \mathbb{R} \\ \mathbf{x} &\mapsto \langle w, \mathbf{x} \rangle \end{aligned}$$

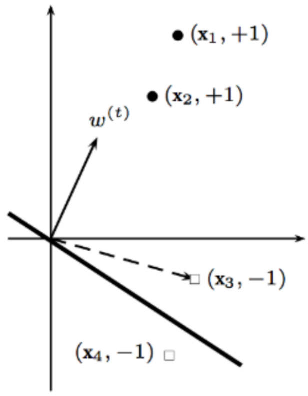
using a training set $S = ((\mathbf{x}_i, y_i))_{i=1}^m$ of size m where, $\langle ., . \rangle$ denotes the dot product and the classes verify $\forall i \in \{1, \ldots, m\}, y_i \in \{-1, +1\}$. The training of the model is generally done on-line as it is shown in algorithm 1.

Algorithm 1 The algorithm of perceptron

```
Training set S = {(x_i, y_i) | i ∈ {1, ..., m}}
Initialize the weights w^(0) ← 0
t ← 0
Learning rate ε > 0
repeat
    Choose randomly an example (x, y) ∈ S
    if y ⟨w^(t), x⟩ < 0 then
        w^(t+1) ← w^(t) + ε × y × x            (A)
        t ← t + 1
    end if
until t > T
```

1. Explain the algorithm.
2. How is called the update rule (eq. (A)), and what does it do?
3. Consider the following classification problem in a two dimensional space.

   Suppose that the chosen example is $\mathbf{x}_3$, what will be the new weight vector using the update rule of the perceptron if $\epsilon = 1$? Draw the weight vector by reproducing the figure in your sheet.



4. We are now interested to demonstrate the convergence of the algorithm in a finite number of iterations and in the case where there exists a weight vector w^* such that $\forall(\mathbf{x}_i, y_i) \in S; y_i \times \langle w^*, \mathbf{x}_i\rangle > 0$. What is the meaning of the condition $y \times \langle w^*, \mathbf{x}\rangle > 0$?

5. We suppose that there exists w^* such that $\forall(\mathbf{x}_i, y_i) \in S; y_i \times \langle w^*, \mathbf{x}_i\rangle > 0$ and we define $\rho = \min_{i\in\{1,\dots,m\}} \left(y_i \langle \frac{w^*}{||w^*||}, \mathbf{x}_i\rangle\right)$. What does ρ represent? Explain why it is a strictly positive real value?

6. We suppose that all the examples in the training set are within a hypersphere of radius R (i.e. $\forall \mathbf{x}_i \in S, ||\mathbf{x}_i|| \leq R$). Further, we initialise the weight vector to be the null vector (i.e. $w^{(0)} = 0$) as well as the learning rate $\epsilon = 1$. Show that after t updates, the norme of the current weight vector satisfies :
$$||w^{(t)}||^2 \leq t \times R^2 \tag{1}$$
hint : You can consider $||w^{(t)}||^2$ as $||w^{(t)} - w^{(0)}||^2$

7. Using the the same condition than in the previous question, show that after t updates of the weight vector we have
$$\left\langle \frac{w^*}{||w^*||}, w^{(t)} \right\rangle \geq t \times \rho \tag{2}$$

8. Deduce from equations (1) and (2) that the number of iterations t is bounded by
$$t \leq \left\lfloor \left(\frac{R}{\rho}\right)^2 \right\rfloor$$
where $\lfloor x \rfloor$ represents the floor function (This result is due to Novikoff, 1966).

9. Explain the previous result.