Perceptron

Carlo Cena

1 Introduction

A perceptron assumes that the data is linearly separable (there exists one hyperplane separating different type of samples without errors). It is therefor able to solve binary problems.

2 Algorithm

Let's define $\boldsymbol{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_n \\ b \end{pmatrix}$ and $\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_m \\ \mathbf{1} \end{pmatrix}$ such that $sign(\boldsymbol{w}^\intercal \boldsymbol{x}_i)$ is equal to the predicted class for $\boldsymbol{x_i}$, consequently a class label will be either 1 or -1.

Therefor, we can say that if $y_i(\boldsymbol{w}^\intercal \boldsymbol{x}_i) > 0$ then features' vector $\boldsymbol{x}_i$ is classified correctly.

The weights' vector $\boldsymbol{w}$ will be updated until all of the x_i are classified correctly ($y_i(\boldsymbol{w}^\intercal \boldsymbol{x}_i) > 0, \forall i$).

2.1 Pseudocode

Set $\boldsymbol{w} = \mathbf{0}$
repeat
    if $y_i(\boldsymbol{w}^\intercal \boldsymbol{x}_i) <= 0$ then
        $\boldsymbol{w} \longleftarrow \boldsymbol{w} + y_i \boldsymbol{x_i}$
    end if
until $y_i(\boldsymbol{w}^\intercal \boldsymbol{x}_i) > 0, \forall i$