Perceptron Convergence Theorem

We will prove the Rosenblatt theorem for convergence of a learning algorithm. If there is a solution, this is, if there is a vector $\mathbf{B}$ that separates a set of vectors in the correct categories with a certain margin k, then in a finite number of steps it is possible to find some vector that also achieves this separation. We call hit when $\mathbf{J}(t) \cdot \boldsymbol{\zeta_\mu} \sigma_\mu > \sqrt{N}k$, otherwise we call a miss.

The vectors of the training set will be considered sequentially and presented to the perceptron, if the classification is correct nothing is done. If an error occurs, a change in the weight vector $\mathbf{J}(t)$ is required. The process is repeated with the next element of list until convergence is achieved. We start from $\mathbf{J}(0) = 0$. The learning dynamics is given by:

$$\mathbf{J}(t+1) = \mathbf{J}(t) + \frac{f_\mu}{\sqrt{N}} \cdot \boldsymbol{\zeta_\mu} \sigma_\mu$$

where f_μ is 0 if we have a hit and 1 otherwise. This is an error correction algorithm.

We suppose that exists $\mathbf{B}$ where $\mathbf{B} \cdot \mathbf{B} = N$ such that:

$$\frac{1}{\sqrt{N}} \boldsymbol{\zeta_\mu} \cdot \mathbf{B}\ \sigma_\mu \geq k > 0 \tag{1.1}$$

At a given time t, we have $F = \sum f_\mu$ and $\mathbf{J}(t) = \sum \frac{1}{\sqrt{N}} f_\mu \boldsymbol{\zeta_\mu} \sigma_\mu$, respectively the number of effective learning steps and the perceptron weights. We want to show that F it remains finite. Multiplying **(1.1)** by f_μ and summing:

$$\sum \frac{1}{\sqrt{N}} f_\mu \boldsymbol{\zeta_\mu} \cdot \mathbf{B}\ \sigma_\mu = \mathbf{J} \cdot \mathbf{B} \geq Fk$$

We introduce $\rho = \frac{\mathbf{J} \cdot \mathbf{B}}{|\mathbf{J}| \cdot |\mathbf{B}|}$. It is easy to show that ρ is between -1 and 1, because is the cosine of the angle between the two vectors. $\mathbf{J} \cdot \mathbf{B} = \rho \sqrt{N} \mid \mathbf{J} \mid$ (because $\mathbf{B} \cdot \mathbf{B} = N$). Squaring the inequality we have:

$$\rho^2 N \mid \mathbf{J} \mid^2 \geq (Fk)^2$$

Analysing the learning dynamic we have:

$$\mid \mathbf{J}(t+1) \mid^2 = \mid \mathbf{J}(t) \mid^2 + 2\frac{1}{\sqrt{N}}\mathbf{J}(t) \cdot \boldsymbol{\zeta}_{\boldsymbol{\mu}}\sigma_\mu + \frac{1}{N} \mid \boldsymbol{\zeta}_{\boldsymbol{\mu}} \mid^2$$

We can use the scale $\mid \boldsymbol{\zeta}_{\boldsymbol{\mu}} \mid^2 = N$ because the classification does not depend of modules.

$$\mid \mathbf{J}(t+1) \mid^2 = \mid \mathbf{J}(t) \mid^2 + 2\frac{1}{\sqrt{N}}\mathbf{J}(t) \cdot \boldsymbol{\zeta}_{\boldsymbol{\mu}}\sigma_\mu + 1$$

Analysing the element μ, if he is changing the dynamic of $\mathbf{J}$ then he was not correct classified. It was a miss, so we have $\frac{1}{\sqrt{N}}\mathbf{J}(t) \cdot \boldsymbol{\zeta}_{\boldsymbol{\mu}}\sigma_\mu < k$ hence:

$$\mid \mathbf{J}(t+1) \mid^2 \leq \mid \mathbf{J}(t) \mid + 2k + 1 \leq F(2k+1)$$

Because every increase step is smaller than 2k+1, so if were given F steps, than $F(2k+1) \geq \mid \mathbf{J}(t) \mid$. Finally we have:

$$(Fk)^2 \leq N \mid \mathbf{J}(t+1) \mid^2 \leq NF(2k+1)$$

$$F \leq N(2k^{-1} + k^{-2})$$

So F is limited.

■