LE PERCEPTRON

NILS HOLZENBERGER

1. Notation

We write $w = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{pmatrix}$ as the weight vector and $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix}$ for each input.

The output of the perceptron is $\text{out}(x_1, ..., x_N) = \begin{cases} 1 & \text{if} \quad w^T x > \theta \\ -1 & \text{else} \end{cases}$

By adding an $N+1-$th coefficient equal to 1 to each x and setting $w_{N+1} = -\theta$:

$\text{out}(x_1, ..., x_{N+1}) = \begin{cases} 1 & \text{if} \quad w^T x > 0 \\ -1 & \text{else} \end{cases}$

We have $p \in \mathbb{N}^*$ inputs which are vectors of $\mathbb{R}^N$ (writing N instead of $N+1$ for convenience), and labels $(y_i)_{1 \leq i \leq p}$

We are looking for w such that $\forall i \in [\![1,p]\!], \begin{cases} w^T x_i > 0 & \text{if} \quad y_i = 1 \\ w^T x_i \leq 0 & \text{if} \quad y_i = -1 \end{cases}$ i.e. $\forall i \in [\![1,p]\!], y_i w^T x_i > 0$ (strict inequalities obtained through continuity, e.g. by tweaking w_N).

Let $T = \begin{pmatrix} y_1 x_1 & y_2 x_2 & ... & y_p x_p \end{pmatrix} \in \mathcal{M}_{N,p}$ ie $T = (y_j x_{j,i})_{i \in [\![1,N]\!], j \in [\![1,p]\!]}$, we are looking for w such that $\forall i \in [\![1,p]\!], [T^T w]_{i,1} > 0$ also written as $T^T w > 0$ and equivalent to $\min\limits_{i \in [\![1,p]\!]} c_i^T w > 0$ where c_i is the $i-$eth column of T

2. Perceptron algorithm

- $w_0 = 0$
- If $T^T w_k > 0$ the algorithm has terminated with solution w_k. Else, let $i \in [\![1,p]\!]$ such that $w_k^T c_i \leq 0$ and

$w_{k+1} = w_k + c_i$

Theorem (Minsky and Papert). The perceptron algorithm finds a solution in a finite number of steps if there exists $u \in \mathbb{R}^N$ such that $T^T u > 0$.

3. Gordan Theorem

Let $A \in \mathcal{M}_{n,p}$, then exactly one of these two equations has a solution:

(i) ${}^tAx > 0$; $x \in \mathcal{M}_{n,1}$

(ii) $x \geq 0$ and $x \neq 0$ and $x \in \text{Ker}A$; $x \in \mathcal{M}_{p,1}$

4. Von Neumann Algorithm

$(e_i)_i$ are the base vectors of $\mathbb{R}^p$

Let $\forall x, y \in \mathbb{R}^p, B(x,y) = {}^tx\, {}^tTTy$ and $q(x) = B(x,x)$.

- $x_0 = e_1$
- Let $j \in [\![1,p]\!]$ such that $\forall i, B(x_k, e_j) \leq B(x_k, e_i)$

$\lambda_k := \arg \min\limits_{\lambda \in [0,1]} q((1-\lambda)x_k + \lambda e_j)$

$x_{k+1} := (1-\lambda_k)x_k + \lambda_k e_j$

Assume (ii) has a solution. Let $\epsilon > 0$. The Von Neumann Algorithm finds, in a finite number of steps, an ϵ-solution, i.e. $x \in \mathbb{R}^p$ such that $||Tx|| \leq \epsilon$

5. Advanced Perceptron

- $w_0 = 0$
- $w_{k+1} = w_k + c_j$ where $j = \arg\min_i {}^t(\frac{w_k}{||w_k||})c_i$ (if $||w_k|| = 0$, stop there, (i) has no solution)

Theorem Assume $\rho(T) > 0$. Let $\epsilon > 0$. The advanced perceptron finds in a finite number of steps $w \in \mathbb{R}^N$ such that $T^T(\frac{w}{||w||}) > (1-\epsilon)\rho(T)$

6. Pál Ruján perceptron

We are trying to find $\rho(T)$ and $w^* := \max_{||w||=1} \min_{i \in [\![1,p]\!]} {}^t w c_i$

Theorem Assume $\rho(T) > 0$. Then w^* belongs to the linear cone generated by $\{c_i\,;\,(w^*)^T c_i = \rho(T)\}$.

The following algorithm is adapted from the algorithm presented in *A Fast Method for Calculating the Perceptron with Maximal Stability* by Pál Ruján.

We build the sequences $A \in (\mathcal{P}([\![1,p]\!]))^{\mathbb{N}}$, and $x \in (\mathbb{R}^n)^{\mathbb{N}}$ such that for all $m \in \mathbb{N}$

(a) $\exists \{\lambda_i\}_{i \in A_m} \subset \mathbb{R}\ ;\ x_m = \sum_{i \in A_m} \lambda_i c_i\ ;\ ||x_m|| = 1$ or $x_m = 0$

(b) $\forall i \in A_m, c_i^T x_m = \alpha_m$

If $x_m = 0$ (or equivalently $\alpha_m = 0$) and $\exists \{\lambda_i\}_{i \in A_m} \subset \mathbb{R}_+^*\ ;\ x_m = \sum_{i \in A_m} \lambda_i c_i$ the algorithm is done.

If $\forall i \in [\![1,p]\!], c_i^T x_m \geq \alpha_m$ and $\exists \{\lambda_i\}_{i \in A_m} \subset \mathbb{R}_+^*\ ;\ x_m = \sum_{i \in A_m} \lambda_i c_i$ the algorithm is done.

- Let $A_0 = \{1\}\ ;\ x_0 = c_1$
- Assume A_m and x_m exist and the algorithm has not terminated at step m

1) Let $c_j \in \{c_i\}_{1 \leq i \leq p}$ such that ${}^t x_m c_j < \alpha_m$

2) Let $S_m = (c_i)_{i \in A_m \cup \{j\}} \in \mathcal{M}_{N,|A_m|+1}$

Let $\begin{pmatrix} z \\ a_1 \\ \vdots \\ a_{|A_m|+1} \end{pmatrix}$ be a solution for $Z \in \mathcal{M}_{|A_m|+2,1}$ in the equation $\left(\begin{array}{c|ccc} 0 & 1 & \cdots & 1 \\ \hline -1 & & & \\ \vdots & & S_m^T S_m & \\ -1 & & & \end{array}\right) Z = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$

If that equation has no solution, consider A'_m obtained from A_m by removing one element, and $S'_m = (c_i)_{i \in A'_m \cup \{j\}}$, and try again.

3) Let $x'_{m+1} = S_m \begin{pmatrix} a_1 \\ \vdots \\ a_{|A_m|+1} \end{pmatrix} \in \mathcal{M}_{N,1}$

If $x'_{m+1} = 0$ then $x_{m+1} = 0$

If $x'_{m+1} \neq 0$ then $x_{m+1} = \frac{x'_{m+1}}{||x'_{m+1}||}$

Then $\alpha_{m+1} = \frac{z}{||x'_{m+1}||}$

We can therefore write $x_{m+1} = \sum_{i \in A_m \cup \{j\}} \mu_i c_i$ where $\mu_i = \frac{a_i}{||x'_{m+1}||}$

Let $k \in A_m$ such that $\mu_k = \min_i \mu_i$

If $\mu_k \leq 0$ then let $A_{m+1} = (A_m \setminus \{k\}) \cup \{j\}$; else $A_{m+1} = A_m \cup \{j\}$

Repeat steps 2 and 3 as long as $\min_i \mu_i \leq 0$