

TD 7 – Learning I Supervised Learning

1 Perceptron model

① Total input

The total input received by the neuron is the weighted sum over all input neurons, which can be written as a scalar product :

$$\sum_j I_{p,j} W_j = \vec{I}_p \cdot \vec{W}$$

② Condition for solving the task

The task is solved if the perceptron is able to categorize each input pattern correctly. This means that the output response r_p is related to the position of the input to the threshold as follows :

$$\forall p, \quad \begin{cases} r_p = 1 \implies \vec{I}_p \cdot \vec{W} > \theta \\ r_p = 0 \implies \vec{I}_p \cdot \vec{W} < \theta \end{cases}$$

③ Rewriting the condition

In order to simplify the condition, the threshold can be modeled as a constant input. To do so, the vectors \vec{I}_p and \vec{W} are appended with a first entry $I_0 = 1$ and $W_0 = -\theta$ respectively, such that :

$$\sum_{j=1}^N I_{p,j} W_j > \theta \iff \sum_{j=1}^N I_{p,j} W_j - \theta \times 1 > 0 \iff \sum_{j=0}^N I_{p,j} W_j > 0$$

The condition becomes :

$$\forall p, \quad \begin{cases} r_p = 1 \implies \vec{I}_p \cdot \vec{W} > 0 \\ r_p = 0 \implies \vec{I}_p \cdot \vec{W} < 0 \end{cases}$$

④ Set of input patterns \vec{J}_p

To reduce the condition to a single equation whose result is always positive, the dot product $\vec{I}_p \cdot \vec{W}$ can be multiplied by a scalar of the same sign, expressed as a function of r_p . The goal is to get a variable s_p such that :

$$\forall p, \quad \begin{cases} r_p = 1 \implies s_p = 1 \\ r_p = 0 \implies s_p = -1 \end{cases}$$

This can be obtained by introducing $s_p = 2r_p - 1$. Then :

$$\forall p, \quad s_p \vec{I}_p \cdot \vec{W} > 0$$

Equivalently, this can be obtained by introducing the vectors $\vec{J}_p = (2r_p - 1)\vec{I}_p$.

$$s_p \vec{I}_p \cdot \vec{W} > 0$$

⑤ Generalization

Generalization is the ability to correctly classify a new input (i.e. not previously seen) after the network has been trained.

2 Perceptron algorithm

⑥ Supervised learning

The correct output is provided by an external "teaching" signal which drives learning in the right direction. The information about the correct output is contained in the vector \vec{J}_p by the appearance of r_p .

⑦ Lower bound

The goal is to express the *numerator* of the cosine between one solution \vec{W}^* vector $\vec{W}(t)$ at a given learning step t . This can be done by recurrence, which requires to express $\vec{W}(t+1) \cdot \vec{W}^*$ as a function of $\vec{W}(t) \cdot \vec{W}^*$:

$$\begin{aligned}\vec{W}(t+1) \cdot \vec{W}^* &= (\vec{W}(t) + \epsilon \vec{J}_p) \cdot \vec{W}^* \\ &= \vec{W}(t) \cdot \vec{W}^* + \epsilon \vec{J}_p \cdot \vec{W}^* \\ &\geq \vec{W}(t) \cdot \vec{W}^* + \epsilon l\end{aligned}$$

This is an arithmetic sequence, which leads by recurrence to :

$$\vec{W}(t) \cdot \vec{W}^* \geq \vec{W}(0) \cdot \vec{W}^* + \epsilon l t = \epsilon l t$$

with $\vec{W}(0) = \vec{0}$.

⑧ Upper bound

The goal is to express the *denominator* of the cosine between one solution \vec{W}^* vector $\vec{W}(t)$ at a given learning step t . This can be done by recurrence, which requires to express $\|\vec{W}(t+1)\|$ as a function of $\|\vec{W}(t)\|$:

$$\begin{aligned}\|\vec{W}(t+1)\|^2 &= \|\vec{W}(t) + \epsilon \vec{J}_p\|^2 \\ &= \|\vec{W}(t)\|^2 + 2\epsilon \vec{J}_p \cdot \vec{W} + \epsilon^2 \|\vec{J}_p\|^2 \\ &\leq \|\vec{W}(t)\|^2 + \epsilon^2 L\end{aligned}$$

Similarly :

$$\|\vec{W}(t)\|^2 \leq \|\vec{W}(0)\|^2 + t\epsilon^2 L = t\epsilon^2 L$$

⑨ Lower bound on $\cos[\alpha(t)]$

Gathering numerator (question ⑦) and denominator (question ⑧) leads to

$$\cos(\alpha(t)) \geq \frac{t\epsilon l}{\sqrt{t\epsilon^2 L}} = \sqrt{t} \frac{l}{L}$$

⑩ Solution

The result obtained at question ⑨ holds *each time a learning step is performed*. In this case, as $\cos(\alpha(t)) < 1$, it imposes that $t < \frac{L^2}{l}$. Thus, the algorithm can only perform a finite number of steps, which means that the algorithm necessarily stops after at most $\frac{L^2}{l}$ steps. When the algorithm stops, all patterns are correctly classified.