

1. The Perceptron Algorithm

$\vec{w}_i = \vec{w}_k$ as stated in the task formulation and assume the training examples $(\vec{x}_{i+1}, y_{i+1}), \dots, (\vec{x}_k, y_k)$ to be already learned. Without loss of generality let

$$y_{i+1}, \dots, y_j \in \{+1\}$$

and

$$y_{j+1}, \dots, y_k \in \{-1\}.$$

Then:

$$\begin{aligned} \vec{w}_k &= \vec{w}_i + \overbrace{\left((\vec{x}_{i+1}, y_{i+1}) + \dots + (\vec{x}_j, y_j) \right) + \left((\vec{x}_{j+1}, y_{j+1}) + \dots + (\vec{x}_k, y_k) \right)}^{=0, \text{ since } \vec{w}_k = \vec{w}_i} \\ &\Rightarrow \vec{x}_{i+1} + \dots + \vec{x}_j = \vec{x}_{j+1} + \dots + \vec{x}_k \end{aligned} \quad (1)$$

Assume \vec{u} to be a solution to the learning problem, i.e. a weight vector for a perceptron that classifies correctly:

$$\langle \vec{u}, \vec{x}_{i+t} \rangle = \begin{cases} \geq 0 & \text{if } t \in \{i+1, \dots, j\} \\ < 0 & \text{if } t \in \{j+1, \dots, k\} \end{cases}$$

Implying

$$\begin{aligned} \langle \vec{u}, \vec{x}_{i+1} + \dots + \vec{x}_j \rangle &\geq 0 \\ \langle \vec{u}, \vec{x}_{j+1} + \dots + \vec{x}_k \rangle &< 0 \end{aligned}$$

in contradiction to (1). Thus the perceptron learning algorithm will loop over the same data infinitely and will never converge. ■

2. Backpropagation with the Hyperbolic Tangent Function

Since $(\tanh(x))' = 1 - \tanh(x)^2$, the weight update rules become:

$$\begin{aligned} \delta_k &\leftarrow (1 - o_k^2) (t_k - o_k) \\ \delta_h &\leftarrow (1 - o_h^2) \sum_{k \in \text{outputs}} w_{h,k} \delta_k \\ w_{i,j} &\leftarrow w_{i,j} + \Delta w_{i,j}, \text{ where } \delta w_{i,j} = \eta \delta_j x_{i,j} \end{aligned}$$

■

3. Backpropagation for the XOR Function

Structure The network has a single hidden layer with three nodes (two regular and one for bias value), a single output node and three input nodes (again two for input and one for a bias value). Thus the structure is the same as seen on slide 27 in Slides from 2014-11-14.

Termination Criterion The termination criterion is on either 100000 iterations or if none of the weights changed. These values can be changed with the command line options `-m` and `-t` respectively.

Example output See the file `output.log`

■

4. Distances

For $\delta(x, y)$ to be a metric, we have to show:

Non-negative $\delta(x, y) \geq 0$

Identity of indiscernibles $\delta(x, y) = 0$ iff $x = y$

Symmetry $\delta(x, y) = \delta(y, x)$

triangle inequality $\delta(x, y) \leq \delta(x, z) + \delta(z, y)$

(i) **Non-negative** because $|\cdot|$ is a norm.

Identity of indiscernibles Clearly the Hamming distance sums to zero iff x_r, x_s don't differ in more than one entry:

$$D_{\text{Hamming}} = 0 \Rightarrow x_r = x_s$$

Symmetry

$$\begin{aligned} D_{\text{Hamming}}(x_r, x_s) &= \sum_{j=1}^m |x_{rj} - x_{sj}| \\ &= \sum_{j=1}^m |x_{sj} - x_{rj}| = D_{\text{Hamming}}(x_s, x_r) \end{aligned}$$

triangle inequality

$$\forall x_r, x_s, x_t :$$

$$\begin{aligned} D_{\text{Hamming}}(x_r, x_t) &= \sum_{j=1}^m |x_{rj} - x_{tj}| \\ &= \sum_{j=1}^m |x_{rj} - x_{sj} + x_{sj} - x_{tj}| \\ &\leq \sum_{j=1}^m (|x_{rj} - x_{sj}| + |x_{sj} - x_{tj}|) \\ &= \sum_{j=1}^m |x_{rj} - x_{sj}| + \sum_{j=1}^m |x_{sj} - x_{tj}| \\ &= D_{\text{Hamming}}(x_r, x_s) + D_{\text{Hamming}}(x_s, x_t) \end{aligned}$$

(ii) **Non-negative** because $|\cdot|$ is a norm.

Identity of indiscernibles

$$D_{\Delta}(S, S) = |S \setminus S \cup S \setminus S| \quad (2)$$

$$= |\emptyset| = 0 \quad (3)$$

and Let $D_{\Delta}(A, B) = 0$ from that follows

$$0 = |A \setminus B \cup B \setminus A| \quad (4)$$

$$= |A \setminus B| + |B \setminus A| \quad \text{because } (A \setminus B) \cap (B \setminus A) = \emptyset \quad (5)$$

and because $|\cdot|$ is positive follows $|A \setminus B| \geq 0$ and $|B \setminus A| \geq 0$ which implies $|A \setminus B| = 0$ and $|B \setminus A| = 0$. From that follows $A \subset B$ and $B \subset A$ which implies $A = B$.

Symmetry $D_{\Delta}(A, B) = |A \setminus B \cup B \setminus A| = |B \setminus A \cup A \setminus B| = D_{\Delta}(B, A)$
because \cup is commutative.

triangle inequality Consider sets A , B and C

$$D_{\Delta}(A, B) + D_{\Delta}(B, C) = |A| + |B| - 2|A \cap B| + |B| + |C| - 2|B \cap C| \quad (6)$$

$$= |A| + |C| + 2(-(|A \cap B| + |B \cap C|) + |B|) \quad (7)$$

$$\geq |A| + |C| + 2(-(|B| + |A \cap B \cap C|) + |B|) \quad (8)$$

$$= |A| + |C| - 2|A \cap B \cap C| \quad (9)$$

$$\geq |A| + |C| - |A \cap C| \quad (10)$$

$$= D_{\Delta}(A, C) \quad (11)$$

■