Group 6: Timm Behner, Philipp Bruckschen, Patrick Kaster, Markus Schwalb
MA-INF 4111 - Intelligent Learning and Analysis Systems: Machine Learning
Exercise Sheet 4

1. The Perceptron Algorithm

Please check my notation!

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

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

and

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

Then:

$$\vec{w}_k = \vec{w}_i + \overbrace{((\vec{x}_{i+1}, y_{i+1}) + \ldots + (\vec{x}_j, y_j)) + ((\vec{x}_{j+1}, y_{j+1}) + \ldots + (\vec{x}_k, y_k))}^{=0,\,\text{since}\,\vec{w}_k=\vec{w}_i}$$

$$\Rightarrow \vec{x}_{i+1} + \ldots + \vec{x}_j = \vec{x}_{j+1} + \ldots + \vec{x}_k \tag{1}$$

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

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

Implying

$$\langle \vec{u}, \vec{x}_{i+1} + \ldots + \vec{x}_j \rangle \geq 0$$
$$\langle \vec{u}, \vec{x}_{j+1} + \ldots + \vec{x}_k \rangle < 0$$

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:

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

∎

3. Backpropagation for the XOR Function

Structure The network has a single hidden layer with two nodes, a single output node and two input nodes.

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 Cleary 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

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

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

Identity of indiscernibles

$$D_\Delta(S,S) = |S \setminus S \cup S \setminus S| \tag{2}$$
$$= |\emptyset| = 0 \tag{3}$$

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

Do we need these equations to be numbered? We need ALL the numbers :D

$$0 = |A \setminus B \cup B \setminus A| \tag{4}$$
$$= |A \setminus B| + |A \setminus B| \qquad \text{because } (A \setminus B) \cap (B \setminus A) = \emptyset \tag{5}$$

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 and distinguish two cases:
Let $x \in A \setminus C$. Then either

(a) $x \notin B \Rightarrow x \in A \setminus B$ or
(b) $x \in B \Rightarrow x \in B \setminus C$.

Similarly: Let $x \in C \setminus A$. Then either

(a) $x \notin B \Rightarrow x \in C \setminus B$ or
(b) $x \in B \Rightarrow x \in B \setminus A$.

$$\begin{aligned} \Rightarrow (A\Delta B) \cup (B\Delta C) &= (A \setminus B) \cup (B \setminus A) \cup (B \setminus C) \cup (C \setminus B) \\ &\supset (A \setminus C) \cup (C \setminus A) \\ &= A\Delta C \end{aligned}$$

$$\Rightarrow |A\Delta C| \leq |(A\Delta B)| + |(B\Delta C)|$$

Alternate version of proof,maybe nicer without the cases

$$\begin{aligned} D_\Delta(A,B) + D_\Delta(B,C) &= |A| + |B| - 2|A \cap B| \\ &\quad + |B| + |C| - 2|B \cap C| && (6) \\ &= |A| + |C| \\ &\quad + 2(-(|A \cap B| + |B \cap C|) + |B|) && (7) \\ &\geq |A| + |C| \\ &\quad + 2(-(|B| + |A \cap B \cap C|) + |B|) && (8) \\ &= |A| + |C| - 2|A \cap B \cap C| && (9) \\ &\geq |A| + |C| - |A \cap C| && (10) \\ &= D_\Delta(A,C) && (11) \end{aligned}$$

■