Ludwig-Maximilians-Universitaet Muenchen 13.05.2015
Institute for Informatics
Prof. Dr. Volker Tresp
Gregor Jossé
Johannes Niedermayer

Machine Learning and Data Mining

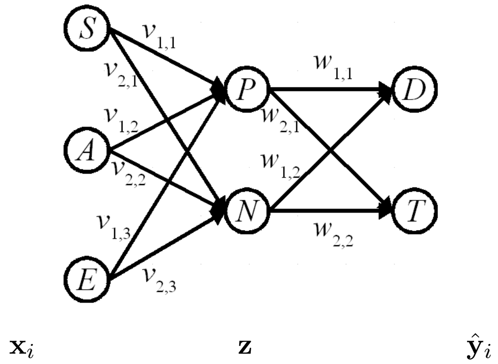
Summer 2015

Exercise Sheet 4

Presentation of Solutions to the Exercise Sheet on the 20.05.2015

Exercise 4-1 Neural Network

Consider the following neural network. It models the game result between two teams D and T, depending on the inputs "self-confidence of team D" (S), "antagonizing power of players of team T" (A) and "efficiency of team D" (E). The hidden neurons model the positive (P) and negative (N) actions of team D.



The output neurons D and T are estimated as follows:

$$\hat{y}_{i,k} = f(\mathbf{x}_i, \mathbf{w}, \mathbf{v})_k = \sum_{h=1}^{M_\phi} w_{k,h}\ \phi_h(\mathbf{x}_i,\mathbf{v}_h)\ ,$$

$$J_N(\mathbf{w}, \mathbf{v}) = \sum_{k=1}^{2} \sum_{i=1}^{N} (y_{i,k} - f(\mathbf{x}_i,\mathbf{w},\mathbf{v})_k)^2\ .$$

The activation function is:

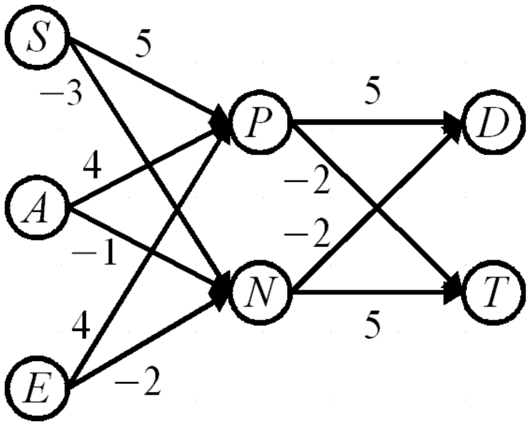
$$z_h(\mathbf{x}_i) = \phi_h(\mathbf{x}_i,\mathbf{v}_h) = \frac{1}{1 + \exp\left(-\sum_{j=1}^{M} v_{h,j} x_{i,j}\right)}\ .$$

The gradient descent for a pattern $\mathbf{x}_i$ is defined as:

$$w_{k,h} \leftarrow w_{k,h} + \eta\ z_h(\mathbf{x}_i)\ (y_{i,k} - f(\mathbf{x}_i, \mathbf{w}, \mathbf{v})_k) \text{ and}$$

$$v_{h,j} \leftarrow v_{h,j} + \eta \sum_{k=1}^{2} w_{k,h}\ z_h(\mathbf{x}_i)\ (1 - z_h(\mathbf{x}_i))\ x_{i,j}\ (y_{i,k} - f(\mathbf{x}_i, \mathbf{w}, \mathbf{v})_k)$$

Consider the already trained neural network:



(a) Compute the prediction $\hat{\mathbf{y}}_i = \begin{pmatrix} D \\ T \end{pmatrix}$ for the input vector $\mathbf{x}_i = \begin{pmatrix} S \\ A \\ E \end{pmatrix} = \begin{pmatrix} -5 \\ 7 \\ 3 \end{pmatrix}$ up to the second decimal.

Possible Solution:

$$\begin{aligned}
P = z_1 = \phi_h(\mathbf{x}_i, \mathbf{v}_1) &= \frac{1}{1 + \exp\left(-\sum_{j=1}^{M} v_{1,j} x_{i,j}\right)} \\
&= \frac{1}{1 + \exp(25 - 28 - 12)} = \frac{1}{1 + \exp(-15)} \approx 1.00 \\
N = z_2 = \phi_h(\mathbf{x}_i, \mathbf{v}_2) &= \frac{1}{1 + \exp(-15 + 7 + 6)} = \frac{1}{1 + \exp(-2)} \approx 0.88 \\
\hat{y}_{i,1} = D = f(\mathbf{x}_i, \mathbf{w}, \mathbf{v})_1 &= \sum_{h=1}^{2} w_{1,h}\, \phi_h(\mathbf{x}_i, \mathbf{v}_h)_1 = 5 \cdot 1 - 2 \cdot 0.88 = 3.24 \\
\hat{y}_{i,2} = T = f(\mathbf{x}_i, \mathbf{w}, \mathbf{v})_2 &= -2 \cdot 1 + 5 \cdot 0.88 = 2.40
\end{aligned}$$

(b) Use the result from (a) and

$$\mathbf{w}_{k,h} \;=\; \mathbf{w}_{k,h} \;+\; \eta \, \frac{\partial J_N(\mathbf{w}, \mathbf{v})}{\partial w_{k,h}}$$

to conduct one part of the update step of the backpropagation algorithm for the value $\mathbf{y}_i = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$. Use a step size of $\eta = 0.5$.

Possible Solution:

$$\begin{aligned}
y_{i,1} - f(\mathbf{x}_i, \mathbf{w}, \mathbf{v})_1 &= 3 - 3.24 = -0.24 \\
y_{i,2} - f(\mathbf{x}_i, \mathbf{w}, \mathbf{v})_2 &= 2 - 2.40 = -0.40 \\
w_{k,h} &\leftarrow w_{k,h} + \eta z_h(\mathbf{x}_i)(y_{i,k} - f(\mathbf{x}_i, \mathbf{w}, \mathbf{v})_k) \\
w_{1,1} &\leftarrow 5 + 0.5 \cdot 1.00 \cdot (-0.24) = 4.88 \\
w_{1,2} &\leftarrow -2 + 0.5 \cdot 0.88 \cdot (-0.24) = -2.1056 \\
w_{2,1} &\leftarrow -2 + 0.5 \cdot 1.00 \cdot (-0.40) = -2.2 \\
w_{2,2} &\leftarrow 5 + 0.5 \cdot 0.88 \cdot (-0.40) = 4.824
\end{aligned}$$