Recurrent perceptron (simplest recurrent NN)

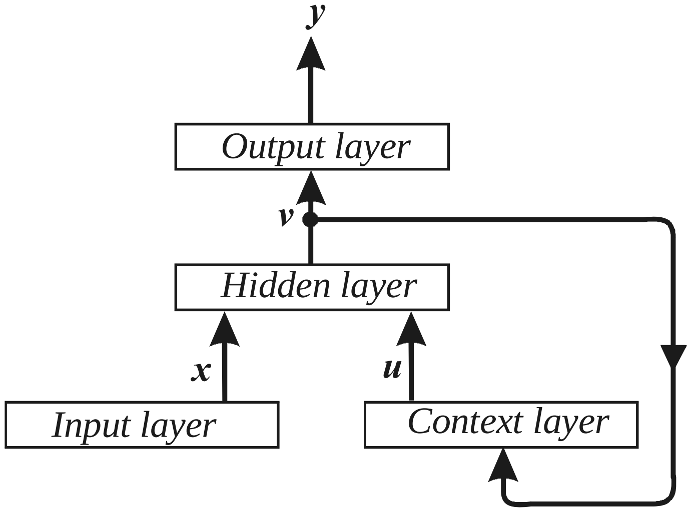
Vladimír Kvasnička

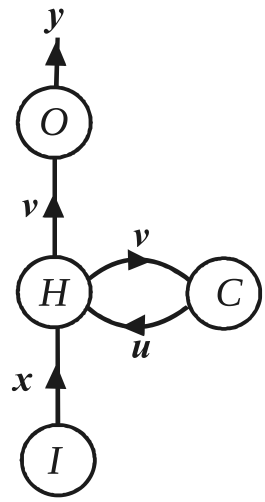
Department of Mathematics
Slovak Technical University
812 37 Bratislava, Slovakia

March 1999

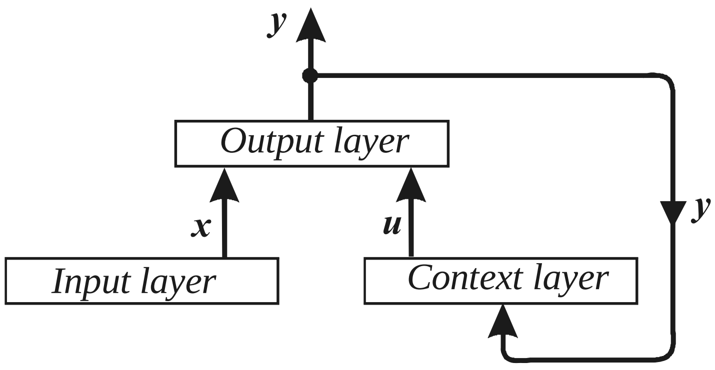
Modified Elman's RNN (recurrent perceptron)

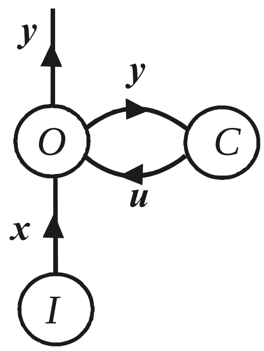
Standard architecture of Elman's recurrent neural network





This architecture is substantially simplified if the hidden neurons are removed, we arrive at the so-called recurrent perceptron (simples recurrent NN)





Activities of neurons are formally determined by a set of equations

$$\left.\begin{array}{l} \boldsymbol{x} = input \in \{a,b,c,d\}^{*} \\ \boldsymbol{u} = C(\boldsymbol{y}) \\ \boldsymbol{y} = O(\boldsymbol{x},\boldsymbol{u}) \in (0,1)^{n} \end{array}\right\} \Rightarrow$$

$$\Rightarrow \begin{cases} \boldsymbol{x} = input \in \{a,b,c,d\}^{*} \\ \boldsymbol{y} = O(\boldsymbol{v}, C(\boldsymbol{y})) \in (0,1)^{n} \quad \text{(a nonlinearity)} \end{cases}$$

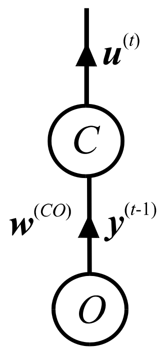
Their iterative solution ($x_{k+1}=f(x_k)$) specifies activities of neurons as follows

$$\left.\begin{array}{l} \boldsymbol{x}^{(t)} = \text{input} \\ \boldsymbol{u}^{(t)} = \begin{cases} 0 & (\text{for } t=1) \\ C\left(\boldsymbol{y}^{(t-1)}\right) & (\text{for } t \geq 2) \end{cases} \\ \boldsymbol{y}^{(t)} = O\left(\boldsymbol{x}^{(t)}, \boldsymbol{u}^{(t)}\right) \end{array}\right\} \quad \text{for } t = 1,2,\ldots,t_{\max}$$

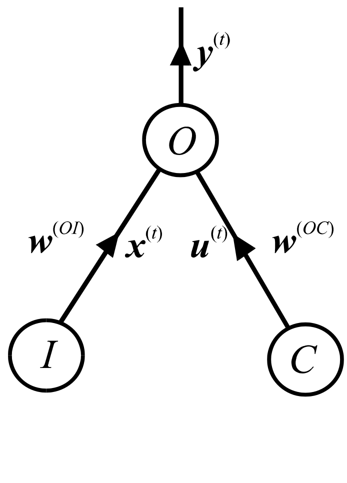
Activities are explicitly determined by

$$x_i^{(t)} = \text{input}$$

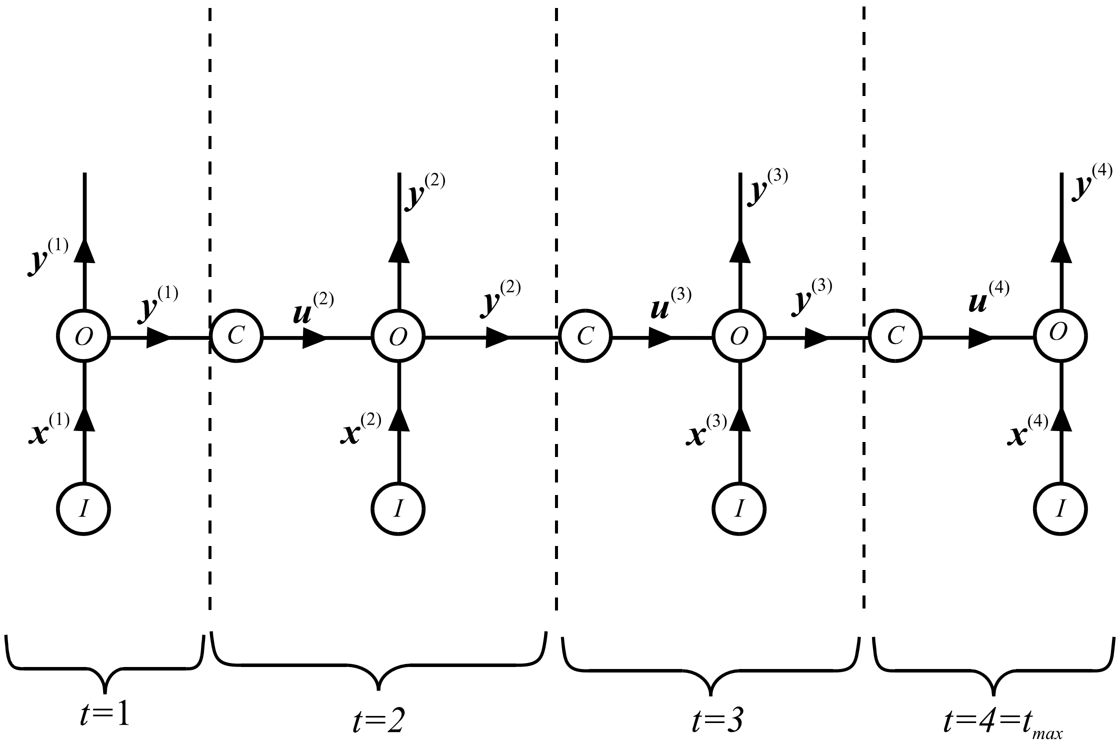
$$u_i^{(t)} = \begin{cases} 0 & (\text{for } t = 1) \\ t\left(\vartheta_i^{(C)} + \sum_{j=1}^{N_O} w_{ij}^{(CO)} y_j^{(t-1)}\right) & (\text{for } t > 1) \end{cases}$$



$$y_i^{(t)} = t\left(\vartheta_i^{(O)} + \sum_{j=1}^{N_I} w_{ij}^{(OI)} x_j^{(t)} + \sum_{j=1}^{N_C} w_{ij}^{(OC)} u_j^{(t)}\right)$$

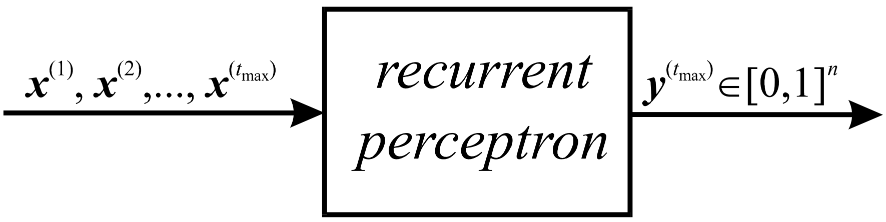


Unfolded recurrent perceptron



$$\mathbf{y}^{(t_{max})} = G\left(\mathbf{x}^{(1)},\mathbf{x}^{(2)},\ldots,\mathbf{x}^{(t_{max})};\mathbf{w},\vartheta\right)$$

The recurrent perceptron can be understood as a parametric mapping of a sequence of inputs $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)},\ldots, \mathbf{x}^{(t_{max})}$ onto a real vector $\mathbf{y}^{(t_{max})} \in [0,1]^n$.



Adaptation (**learning**) of the recurrent neural network

$$A_{train} = \left\{ \left(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(t_{max})} \right) / \left(\boldsymbol{y}_{req} \right) \right\}$$

$$E = \frac{1}{2} \left(\boldsymbol{y}^{(t_{max})} - \boldsymbol{y}_{req} \right)^2$$

An adaptation is equivalent to a minimization of the objective function E with respect to weight and threshold coefficients

$$\boxed{\left(w_{opt}, \vartheta_{opt} \right) = \arg \min_{(w,\vartheta)} E(w,\vartheta)}$$

This optimization problem is most frequently solved by the so-called **steepest descent** gradient method

$$w_{k+1} = w_k - \lambda \, grad \; E(\boldsymbol{x}_k)$$

In general, partial derivatives of the objective function E are determined as follows

$$\left(\frac{\partial E}{\partial \vartheta_i}\right) = t'(\xi_i)\left(g_i + \sum_k \frac{\partial E}{\partial \vartheta_k} w_{ki}\right)$$

$$t'(\xi_i) = t(\xi_i)\left[1 - t(\xi_i)\right]$$

$$g_i = \begin{cases} x_i - x_{i,req} & (i \in O) \\ 0 & (i \notin O) \end{cases}$$

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial \vartheta_i} x_j$$

These general formulae (they form a background of the so-called **back-propagation approach**) are immediately applicable to unfolded recurrent neural networks for calculation of partial derivatives of the objective function E.

(1) Initialization, t=$t_{\max}$

$$\left(\frac{\partial E}{\partial \vartheta_i^{(O)}}\right)^{(t_{max})} = y_i^{(t_{max})}\left(1-y_i^{(t_{max})}\right)\left(y_i^{(t_{max})}-y_{i,req}\right)$$

$$\left(\frac{\partial E}{\partial \vartheta_i^{(C)}}\right)^{(t_{max})} = u_i^{(t_{max})}\left(1-u_i^{(t_{max})}\right)\sum_{j=1}^{N_O}\left(\frac{\partial E}{\partial \vartheta_j^{(O)}}\right)^{(t_{max})} w_{ji}^{(OC)}$$

(2) Iteration, $1 \le t < t_{\max}$

$$\left(\frac{\partial E}{\partial \vartheta_i^{(O)}}\right)^{(t)} = y_i^{(t)}\left(1-y_i^{(t)}\right)\sum_{j=1}^{N_C}\left(\frac{\partial E}{\partial \vartheta_j^{(C)}}\right)^{(t+1)} w_{ji}^{(CO)}$$

$$\left(\frac{\partial E}{\partial \vartheta_i^{(C)}}\right)^{(t)} = u_i^{(t)}\left(1-u_i^{(t)}\right)\sum_{j=1}^{N_O}\left(\frac{\partial E}{\partial \vartheta_j^{(O)}}\right)^{(t)} w_{ji}^{(OC)}$$

Partial derivatives with respect to weight coefficients are determined by

(1) Initialization, $t=t_{\max}$

$$\left(\frac{\partial E}{\partial w_{ij}^{(OI)}}\right)^{(t_{max})}=\left(\frac{\partial E}{\partial \vartheta_i^{(O)}}\right)^{(t_{max})}x_j^{(t_{max})}$$

$$\left(\frac{\partial E}{\partial w_{ij}^{(OC)}}\right)^{(t_{max})}=\left(\frac{\partial E}{\partial \vartheta_i^{(O)}}\right)^{(t_{max})}u_j^{(t_{max})}$$

(2) Iteration, $1\leq t<t_{\max}$

$$\left(\frac{\partial E}{\partial w_{ij}^{(CO)}}\right)^{(t)}=\left(\frac{\partial E}{\partial \vartheta_i^{(C)}}\right)^{(t+1)}y_j^{(t)}$$

$$\left(\frac{\partial E}{\partial w_{ij}^{(OC)}}\right)^{(t)}=\left(\frac{\partial E}{\partial \vartheta_i^{(O)}}\right)^{(t)}u_j^{(t)}$$

$$\left(\frac{\partial E}{\partial w_{ij}^{(OI)}}\right)^{(t)}=\left(\frac{\partial E}{\partial \vartheta_i^{(O)}}\right)^{(t)}x_j^{(t)}$$

Total partial derivatives are determined as follows

$$\frac{\partial E}{\partial \vartheta_i^{(O)}} = \sum_{t=1}^{t_{max}} \left(\frac{\partial E}{\partial \vartheta_i^{(O)}} \right)^{(t)}$$

$$\frac{\partial E}{\partial \vartheta_i^{(C)}} = \sum_{t=2}^{t_{max}} \left(\frac{\partial E}{\partial \vartheta_i^{(C)}} \right)^{(t)}$$

$$\frac{\partial E}{\partial w_{ij}^{(OC)}} = \sum_{t=2}^{t_{max}} \left(\frac{\partial E}{\partial w_{ij}^{(OC)}} \right)^{(t)} = \sum_{t=2}^{t_{max}} \left(\frac{\partial E}{\partial \vartheta_i^{(O)}} \right)^{(t)} u_j^{(t)}$$

$$\frac{\partial E}{\partial w_{ij}^{(CO)}} = \sum_{t=1}^{t_{max}-1} \left(\frac{\partial E}{\partial w_{ij}^{(CO)}} \right)^{(t)} = \sum_{t=1}^{t_{max}-1} \left(\frac{\partial E}{\partial \vartheta_i^{(C)}} \right)^{(t+1)} y_j^{(t)}$$

$$\frac{\partial E}{\partial w_{ij}^{(OI)}} = \sum_{t=1}^{t_{max}} \left(\frac{\partial E}{\partial w_{ij}^{(OI)}} \right)^{(t)} = \sum_{t=1}^{t_{max}} \left(\frac{\partial E}{\partial \vartheta_i^{(O)}} \right)^{(t)} x_j^{(t)}$$