2. prednáška

Logické neuróny a neurónové siete

Mozog a neurónové siete

- ***Metafora ľudského mozgu*** hrá dôležitú úlohu v modernej informatike. Pomocou tejto metafory boli navrhnuté nové paralelné prístupy k spracovaniu informácie, ktoré sa zásadným spôsobom odlišujú od klasickej sekvenčnej architektúry počítačov podľa von Neumanna.

- Tento nový prístup k spracovaniu informácie je reprezentovaný ***teóriou umelých neurónových sietí***, ktorá v súčasnosti tvorí nielen efektívny informatický nástroj na tvorbu a návrh nových paralelných prístupov k riešeniu problémov umelej inteligencie, ale už je aj ***integrálnou súčasťou modernej neurovedy***, pomocou ktorej sa pristupuje k počítačovým simuláciám procesov prebiehajúcich v mozgu.

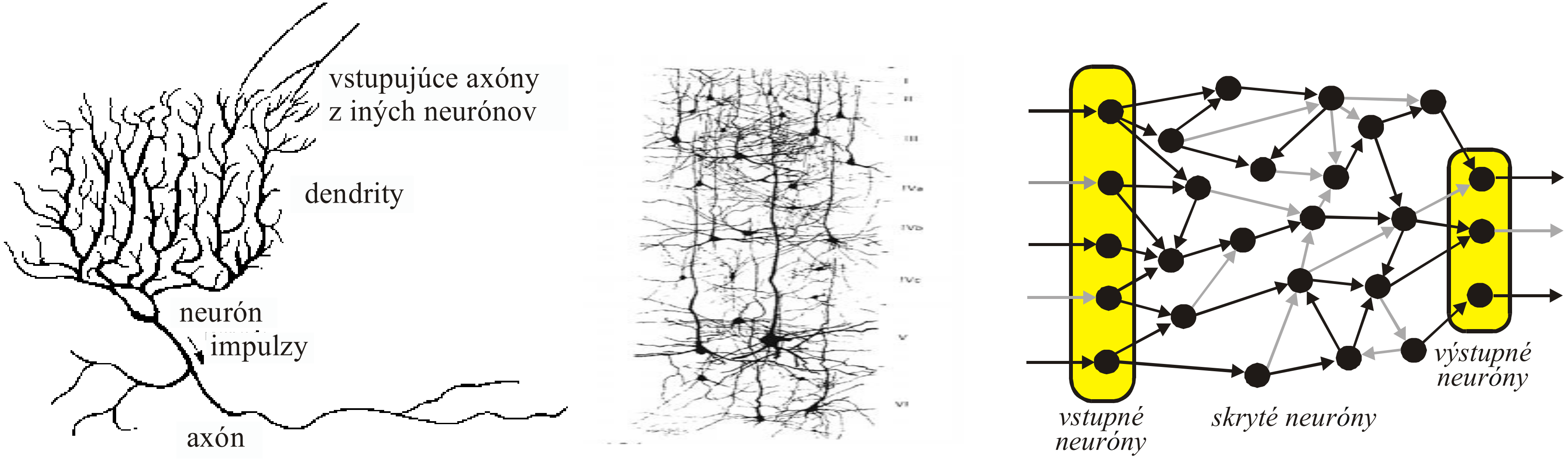
Neurónové siete sú založené na základnom postuláte neurovedy, podľa ktorého základným stavebným kameňom ľudského mozgu je neurón, ktorý má tieto tri najdôležitejšie vlastnosti:

(1) neurón má orgán, pomocou ktorého ***prijíma signály*** z okolia od ostatných neurónov (dendrity),

(2) neurón ***spracováva*** (*integruje*) prijaté signály,

(3) neurón má orgán (axón) pomocou ktorého ***posiela spracované vstupné signály iným neurónom zo svojho okolia***.

Neuróny sú poprepájané medzi sebou do zložitej sieťovej štruktúry (nazývanej neurónová sieť), pričom jednotlivé ***spoje majú buď excitačný alebo inhibičný charakter***. Systém spojov a ich excitačný resp. inhibičný charakter tvoria ***architektúru neurónovej siete***, ktorá jediná určuje vlastnosti neurónovej siete.



A B C

(A) Typická neurónová bunka, ktorá obsahuje rozsiahly dendritický systém (vstup) a dlhý vetviaci sa axón (výstup).
(B) Vzájomné prepojenie neurónov pomocou spojov medzi dendritmi a axónmi, ktoré je pozorované v mikroanatómii ľudského mozgu.
(C) Architektúra neurónovej siete s dvoma druhmi spojov medzi neurónmi, tmavé (svetlé) spoje znázorňujú excitačné (inhibičné) spoje.

Logické neuróny a teória neurónových sietí

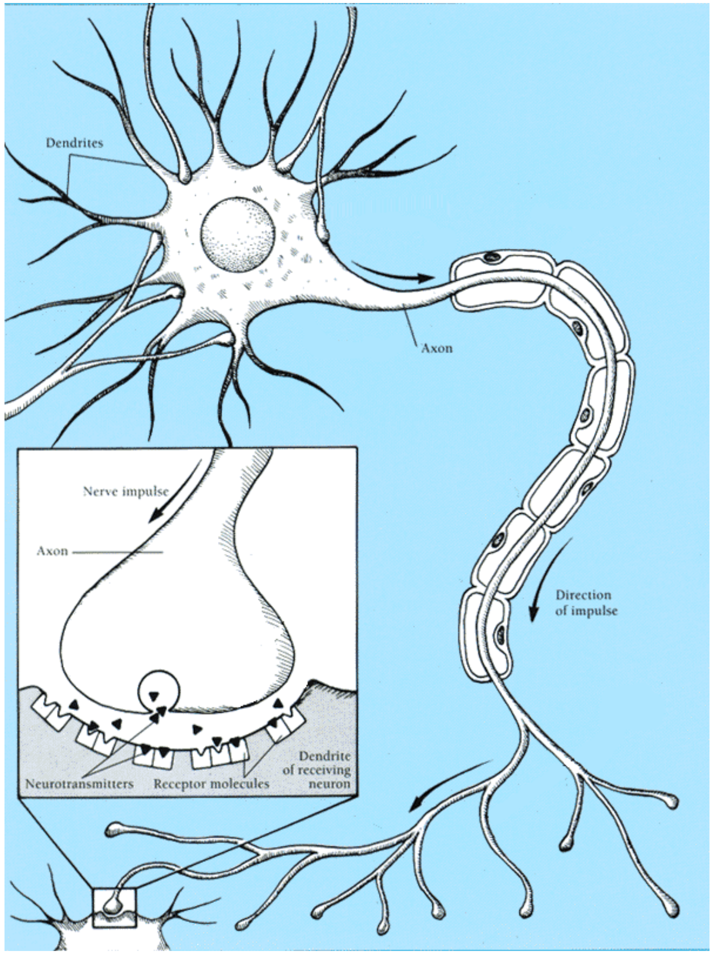
- Budeme študovať najjednoduchší typ neurónových sietí, ktoré boli navrhnuté už pred viac ako polstoročím ***McCullochom a Pittsom*** v r. 1943. Ich model neurónovej siete je významný medzník v rozvoji teórie neurónových sietí.

- Elementárnou jednotkou McCullochovej a Pittsovej neurónovej siete je ***logický neurón*** (výpočtová jednotka), pričom stav neurónu je binárny (tj. má dva možné stavy, 1 a 0).

- Dendritický systém logického neurónu obsahuje tak *excitačné vstupy* (opísané binárnymi premennými x_1, x_2, ..., x_n, ktoré zosilňujú odozvu), ako aj *inhibičné vstupy* (opísané binárnymi premennými x_{n+1}, x_{n+2}, ..., x_m, ktoré zoslabujú odozvu).

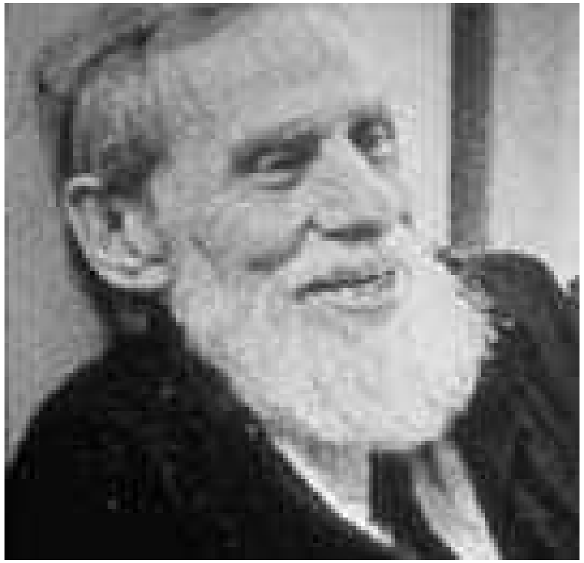
Základné skutočnosti z neurovedy

Podrobná schéma neurónu



obrázok je prevzatý z
http://www.pfizer.com/brain/image

Budeme študovať jeden z najjednoduchších modelov neurónov, ktorý v r. 1943 navrhli McCulloch a Pitts. Ich jednoduchý model logického neurónu je v súčasnosti pokladaný za významný medzník v rozvoji umelej inteligencie (menovite teórie umelých neurónových sietí).



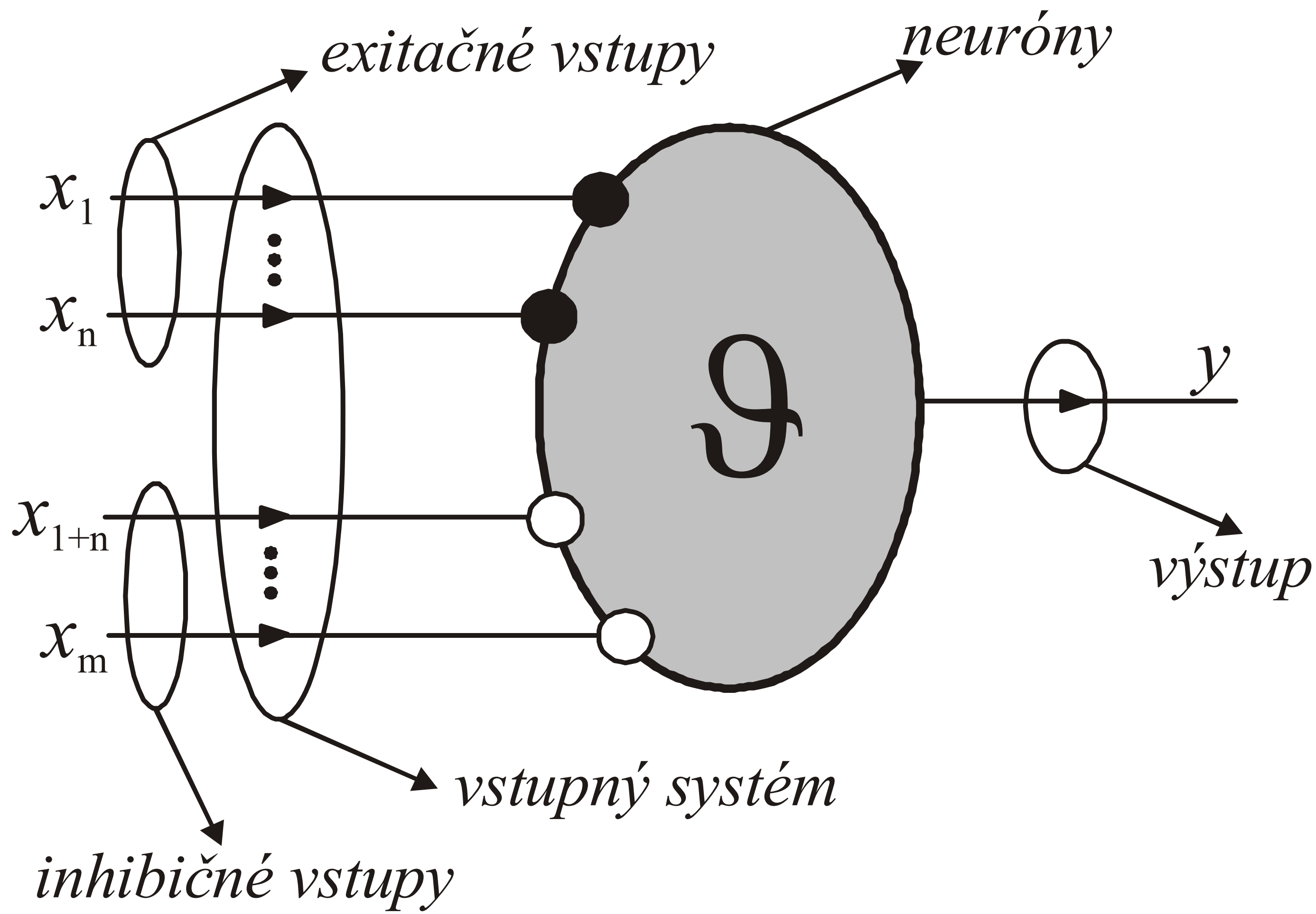
Waren McCulloch (1898-1972)

Walter Pitts (1923-196?)

Neurónové siete sú založené na postuláte neurovedy hovoriacom, že základným stavebným kameňom ľudského mozgu je ***neurón***, ktorý má tieto tri najdôležitejšie vlastnosti:

- neurón má orgán, pomocou ktorého *prijíma signály* z okolia od ostatných neurónov (dendrity),
- neurón *spracováva* (*integruje*) prijaté signály,
- neurón má orgán (axón) pomocou ktorého *posiela spracované vstupné signály iným neurónom zo svojho okolia*.

Logický neurón McCullocha a Pittsa

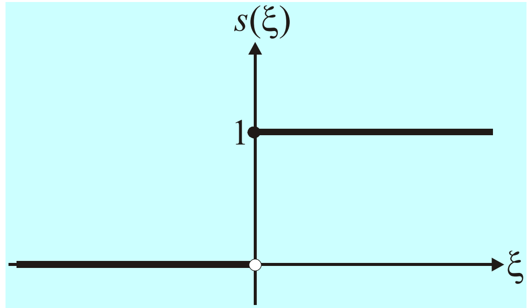


Logický neurón je založený na pravidle:

aktivita je jednotková, ak *vnútorný potenciál* neurónu definovaný ako rozdiel medzi sumou excitačných vstupných aktivít a inhibičných vstupných aktivít je väčší alebo rovný prahu ϑ, v opačnom prípade je nulová

$$y = \begin{cases} 1 & \left(\xi = x_1 + \ldots + x_n - x_{1+n} - \ldots - x_m \geq \vartheta\right) \\ 0 & \left(\xi < \vartheta\right) \end{cases}$$

Použitím jednoduchej „schodovej“ funkcie



$$s(\xi)=\begin{cases}1 & (\text{ak } \xi \geq 0)\\ 0 & (\text{ak } \xi < 0)\end{cases}$$

$$y=s\left(\underbrace{x_1+\ldots+x_n-x_{1+n}-\ldots-x_m}_{\xi}-\vartheta\right)$$

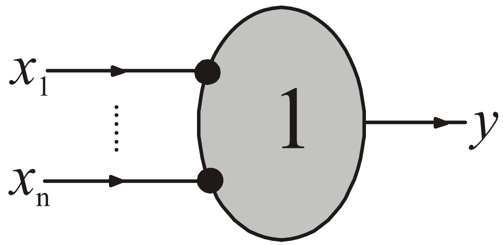
Aktivita logického neurónu môže byť alternatívne vyjadrená pomocou binárnych váhových koeficientov

$$\boxed{y=s\left(\underbrace{w_1x_1+\ldots+w_mx_m}_{\xi}-\vartheta\right)=s\left(\sum_{i=1}^{m}w_ix_i-\vartheta\right)}$$

Implementácia Boolovej binárnej funkcie OR

$$y_{OR}\left(x_1, x_2\right) = s\left(x_1 + x_2 - 1\right)$$

#	x_1	x_2	$y_{OR}(x_1,x_2)$	$x_1 \vee x_2$
1	0	0	$s(-1)$	0
2	0	1	$s(0)$	1
3	1	0	$s(0)$	1
4	1	1	$s(1)$	1



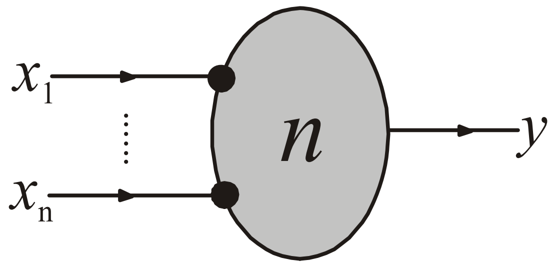
Boolova funkcia disjunkcie

$$y = x_1 \vee \ldots \vee x_n$$

Implementácia Boolovej binárnej funkcie AND

$$y_{AND}\left(x_1, x_2\right) = s\left(x_1 + x_2 - 2\right)$$

#	x_1	x_2	$y_{AND}(x_1,x_2)$	$x_1 \wedge x_2$
1	0	0	$s(-2)$	0
2	0	1	$s(-1)$	0
3	1	0	$s(-1)$	0
4	1	1	$s(0)$	1



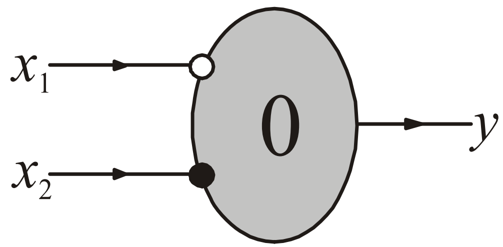
Boolova funkcia konjunkcie

$$y = x_1 \wedge \ldots \wedge x_n$$

Implementácia Boolovej binárnej funkcie IF-THEN

$$y_{AND}\left(x_1,x_2\right)=s\left(-x_1+x_2\right)$$

#	x_1	x_2	$y_{IF\text{-}THEN}(x_1,x_2)$	$x_1 \Rightarrow x_2$
1	0	0	$s(0)$	1
2	0	1	$s(1)$	1
3	1	0	$s(-1)$	0
4	1	1	$s(0)$	1

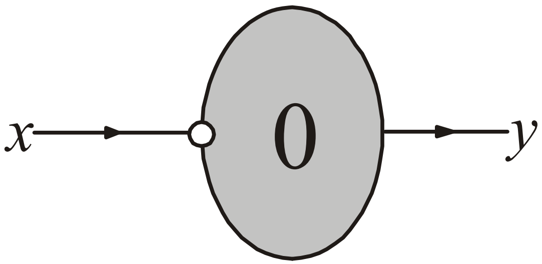


Boolova funkcia implikácie
$y = x_1 \Rightarrow x_2$

Implementácia Boolovej inárnej funkcie NOT

$$y_{NOT}\left(x\right) = s\left(-x+0\right)$$

#	x	$y_{NOT}(x)$	$\neg x$
1	0	$s(0)$	1
2	1	$s(-1)$	0

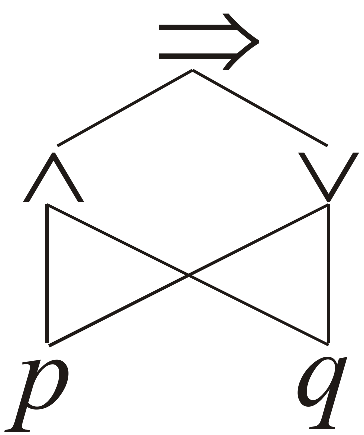


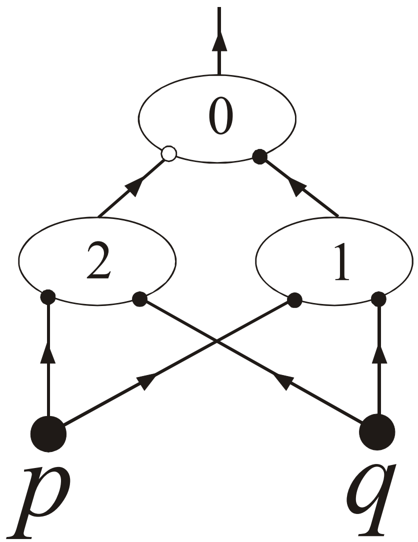
Boolova funkcia negácie
$y = \neg x$

Veta.

Každá Boolova funkcia môže byť vyjadrená pomocou „neurónovej siete“ zloženej z logických neurónov vyjadrujúcich logické spojky.

$$\varphi(p,q)=(p\wedge q)\Rightarrow(p\vee q)$$

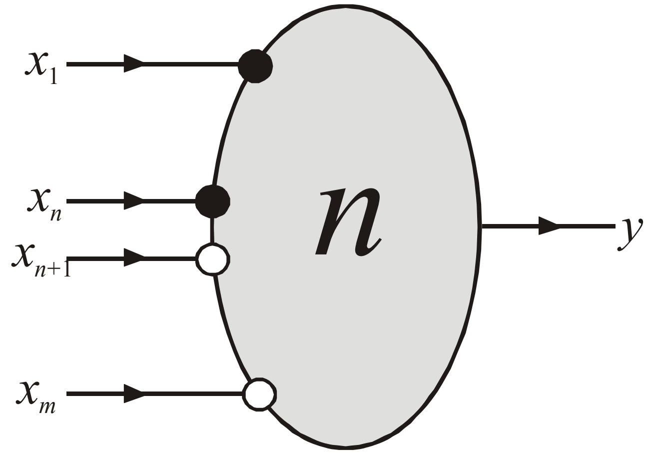




Implementácia konjunktívej klauzule

$$x_1 \wedge \ldots \wedge x_n \wedge \neg x_{n+1} \wedge \ldots \wedge \neg x_m$$

$$val_\tau\left(x_1 \wedge \ldots \wedge x_n \wedge \neg x_{n+1} \wedge \ldots \wedge \neg x_m\right) = \begin{cases} 1 & \left(x_1 = \ldots = x_n = 1\ a\ x_{n+1} = \ldots = x_m = 0\right) \\ 0 & \left(in\acute{a}\check{c}\right) \end{cases}$$



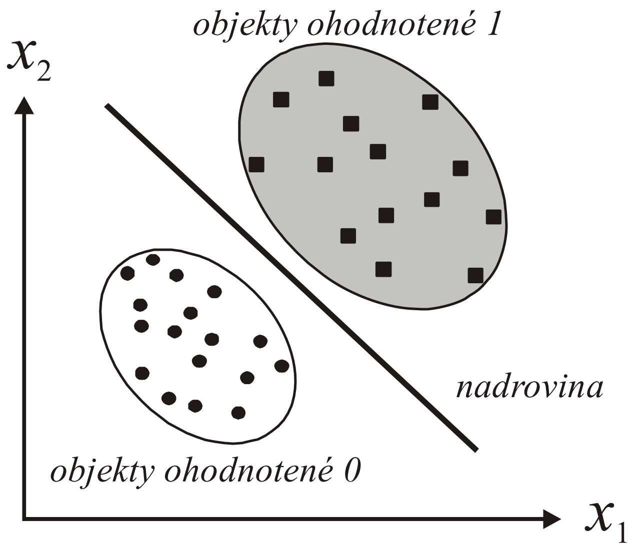
$$y = s\left(x_1 + \ldots + x_n - x_{n+1} - \ldots - x_m - n\right)$$

Lineárna separovateľnosť

Geometrická interpretácia výpočtu prebiehajúceho v logickom neuróne:

$$w_1x_1 + w_2x_2 + ... + w_nx_n - \vartheta = 0$$

Táto rovina rozdeľuje stavový priestor na dva polopriestory.



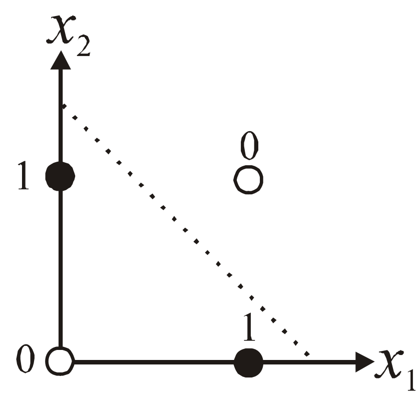
Definícia.
Boolova funkcia $f(x_1, x_2,..., x_n)$ je *lineárne separovateľná*, ak existuje taká rovina $w_1x_1 + w_2x_2 + ...+ w_nx_n - \vartheta = 0$, ktorá separuje priestor vstupných aktivít tak, že v jednej časti priestoru sú vrcholy ohodnotené 0, zatiaľ čo v druhej časti priestoru sú vrcholy ohodnotené 1.

Veta.
Logický neurón je schopný simulovať len tie Boolove funkcie, ktoré sú lineárne separovateľné.

Boolovej funkcie XOR nie je lineárne separovateľná

$$\varphi_{XOR}\left(x,y\right)=x\oplus y$$

#	x	y	$\varphi_{XOR}(x,y)$
1	0	0	0
2	0	1	1
3	1	0	1
4	1	1	0



V teórii Boolovych funkcií sa dokázuje veta, že každá výroková formula môže byť prepísaná do ekvivalentného disjunktívneho tvaru

$$\varphi = \bigvee_{\substack{\tau \\ (val_\tau(\varphi)=1)}} x_1^{(\tau)} \wedge x_2^{(\tau)} \wedge \ldots \wedge x_n^{(\tau)}$$

kde

$$x_i^{(\tau)} = \begin{cases} x_i & \left(\text{ak } val_\tau(x_i) = 1\right) \\ \neg x_i & \left(\text{ak } val_\tau(x_i) = 0\right) \end{cases}$$

Ilustračný príklad

#	x_1	x_2	x_3	$y = f(x_1, x_2, x_3)$	*konjunktívna klauzula*
1	0	0	0	0	
2	0	0	1	0	
3	0	1	0	1	$\overline{x}_1 \wedge x_2 \wedge \overline{x}_3$
4	0	1	1	1	$\overline{x}_1 \wedge x_2 \wedge x_3$
5	1	0	0	0	
6	1	0	1	1	$x_1 \wedge \overline{x}_2 \wedge x_3$
7	1	1	0	0	
8	1	1	1	0	

$$y = f(x_1, x_2, x_3) = (\overline{x}_1 \wedge x_2 \wedge \overline{x}_3) \vee (\overline{x}_1 \wedge x_2 \wedge x_3) \vee (x_1 \wedge \overline{x}_2 \wedge x_3)$$

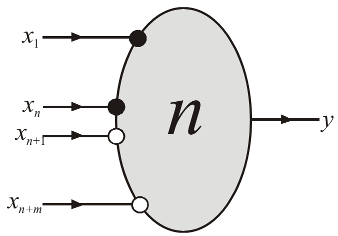
Táto Boolova funkcia môže byť zjednodušená tak, že prvá a druhá klauzula sa zjednodušia

$$\left(\overline{x}_1 \wedge x_2 \wedge \overline{x}_3\right) \vee \left(\overline{x}_1 \wedge x_2 \wedge x_3\right) = \overline{x}_1 \wedge x_2 \wedge \underbrace{\left(\overline{x}_3 \vee x_3\right)}_{1} = \overline{x}_1 \wedge x_2$$

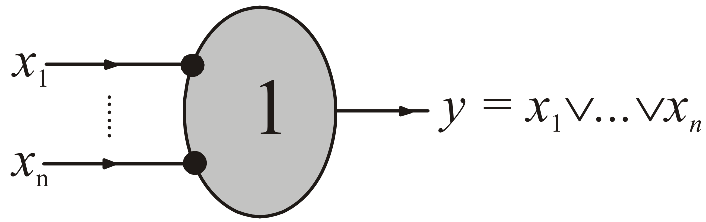
Potom

$$y = f\left(x_1, x_2, x_3\right) = \left(\overline{x}_1 \wedge x_2\right) \vee \left(x_1 \wedge \overline{x}_2 \wedge x_3\right)$$

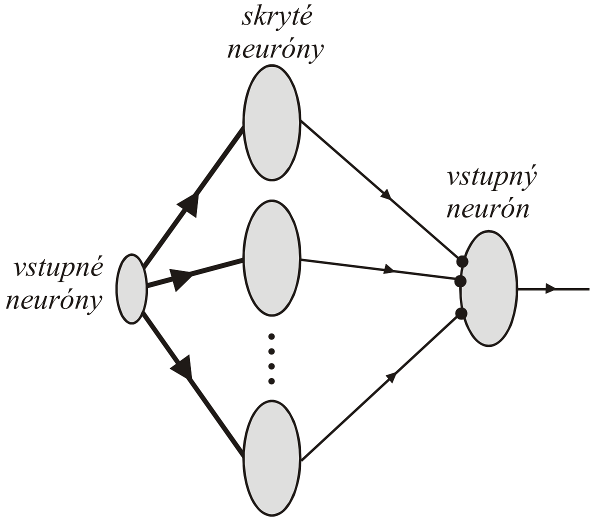
Konjunktívne klauzule $x_1^{(\tau)} \wedge x_2^{(\tau)} \wedge \ldots \wedge x_n^{(\tau)}$ môžeme vyjadriť jedným logickým neurónom



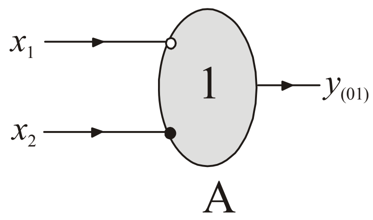
Výstupy z týchto neurónov spojíme do disjunkcie pomocou neurónu



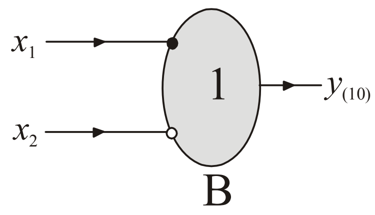
Trojvrstvová neurónová sieť



Ilustračný príklad – konštrukcia XOR funkcie

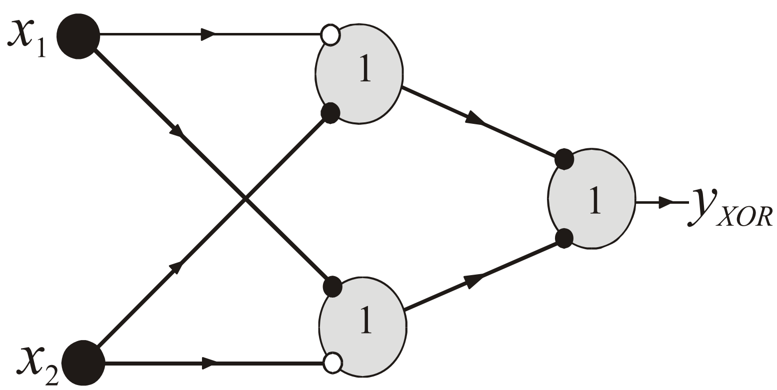


$\neg x_1 \wedge x_2$



$x_1 \wedge \neg x_2$

$$\varphi_{XOR}\left(x_1, x_2\right) = \left(\neg x_1 \wedge x_2\right) \vee \left(x_1 \wedge \neg x_2\right)$$



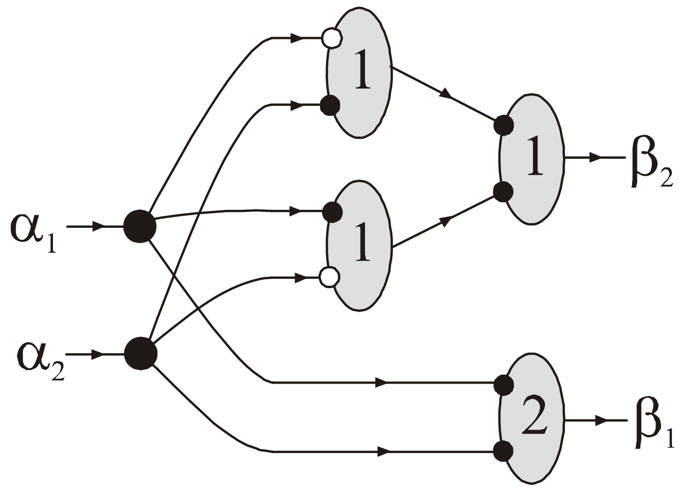
Veta.
Ľubovoľná Boolova funkcia f je simulovaná pomocou 3-vrstvovej neurónovej siete.

Komentár.
3-vrstvové neurónové siete obsahujúce logické neuróny sú *univerzálne výpočtové zariadenia* pre doménu Boolových funkcií.

Príklad

$$\frac{\begin{array}{c}\alpha_1\\ \alpha_2\end{array}}{\beta_1\beta_2}$$

$$\beta_1 = \alpha_1 \wedge \alpha_2$$
$$\beta_2 = \alpha_1 \oplus \alpha_2 = \left(\neg\alpha_1 \wedge \alpha_2\right) \vee \left(\alpha_1 \wedge \neg\alpha_2\right)$$



Logické neuróny vyšších rádov

Ak vnútorný potenciál neurónu ξ je určený ako lineárna kombinácia vstupných aktivít (t.j. len prvou sumou), potom logický neurón je štandardný a nazýva sa "***logický neurón prvého rádu***". Ak tento potenciál neurónu ξ obsahuje aj kvadratické a prípadne aj ďalšie členy, potom sa nazýva "***logický neurón vyššieho rádu***".

$$y = s\left(\underbrace{\sum_{i=1}^{n} w_i x_i + \sum_{\substack{i,j=1 \\ (i<j)}}^{n} w_{ij} x_i x_j + \ldots - \vartheta}_{\xi} \right)$$

Podľa Minského a Paperta (v knihe *Perceptron*) perceptróny vyššieho rádu sú schopné simulovať aj množiny objektov, ktoré nie sú lineárne separovateľné.

Veta.
Ľubovolná Boolova funkcia f je Boolova funkcia je simulovaná logickým neurónom vyššieho rádu.

Príklad

Boolovu binárnu funkciu *XOR* vyjadríme pomocou logického neurónu druhého rádu

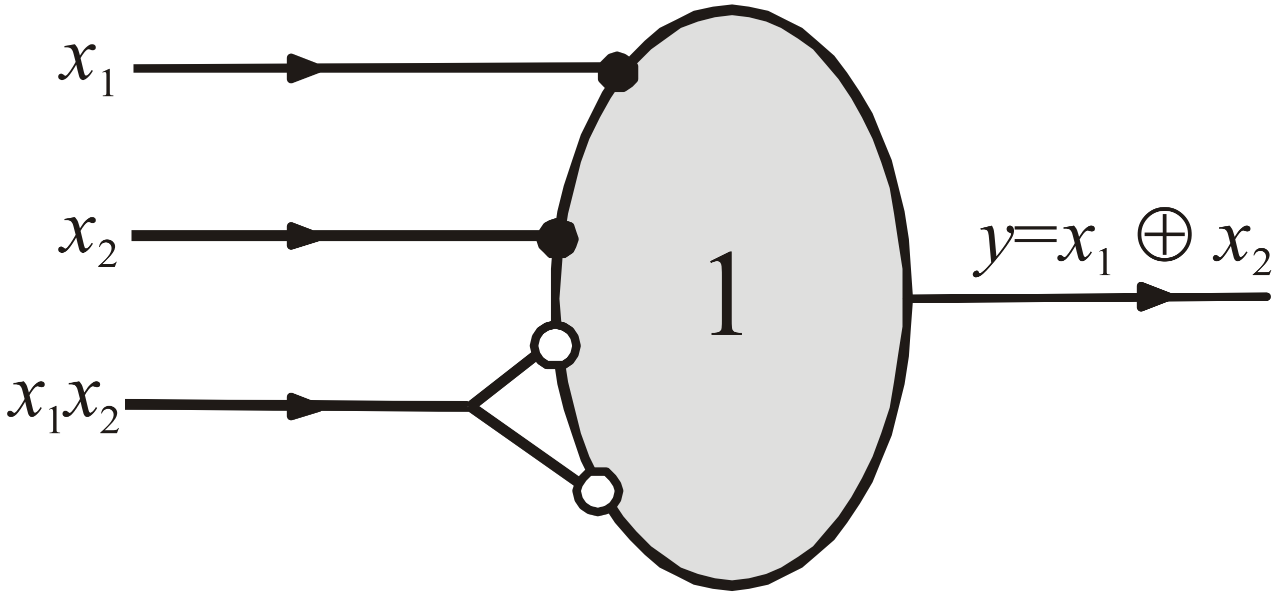
$$y = s\left(\underbrace{w_1 x_1 + w_2 x_2 + w_{12} x_1 x_2}_{\xi} - \vartheta\right)$$

Koeficienty sú určené podmienkami

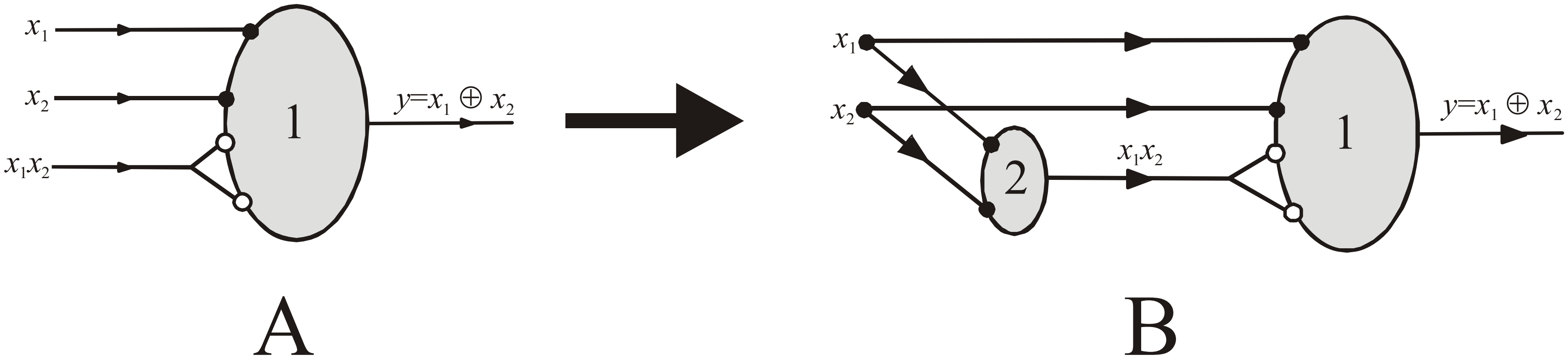
$$\begin{aligned}
-\vartheta &< 0 \quad \left(pre\,(0,0)/0\right)\\
w_2 \qquad -\vartheta &\geq 0 \quad \left(pre\,(0,1)/1\right)\\
w_1 \qquad\qquad -\vartheta &\geq 0 \quad \left(pre\,(1,0)/1\right)\\
w_1 + w_2 + w_{12} - \vartheta &< 0 \quad \left(pre\,(1,1)/0\right)
\end{aligned}$$

Postupným riešením týchto nerovníc dostaneme, že váhové koeficienty a prah majú napr. takéto riešenie

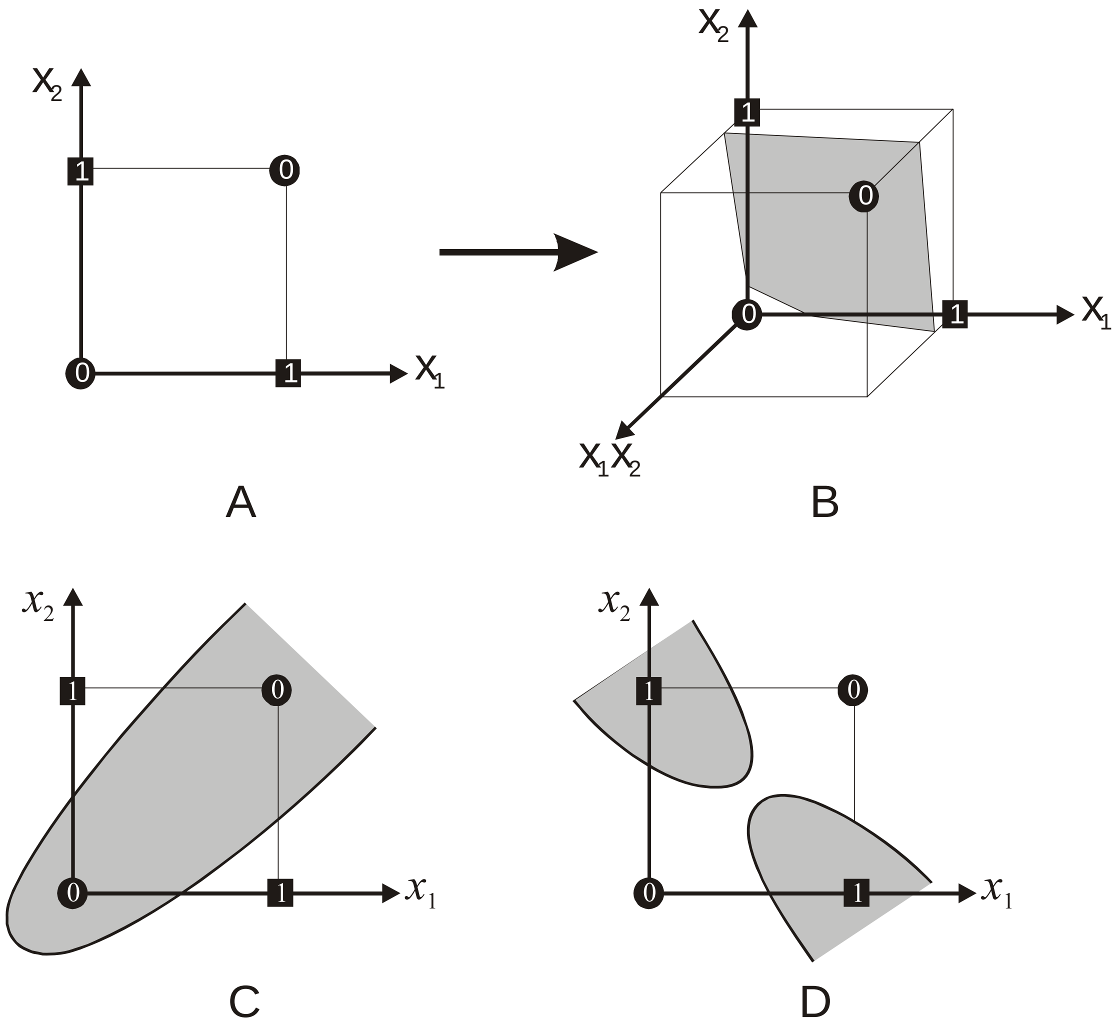
$$\vartheta = 1,\ w_1 = w_2 = 1,\ w_{12} = -2$$



Transformácia logického neurónu 2. rádu na neurónovú sieť obsahujúcu dva logické neuróny prvého rádu



Grafická interpretácia logického neurónu 2. rádu



Definícia.
Boolova funkcia f sa nazýva *kvadraticky separovateľná*, ak existujú také váhové koeficienty w_i, w_{ij} a prahový faktor ϑ, že pre každú špecifikáciu premenných $x_1, x_2, ..., x_n$ platí

$$y_{req}\left(x_1, x_2, ..., x_n\right) = 1 \Rightarrow \sum_{i=1}^{n} w_i x_i + \sum_{\substack{i,j=1 \\ (i<j)}}^{n} w_{ij} x_i x_j \geq \vartheta$$

$$y_{req}\left(x_1, x_2, ..., x_n\right) = 0 \Rightarrow \sum_{i=1}^{n} w_i x_i + \sum_{\substack{i,j=1 \\ (i<j)}}^{n} w_{ij} x_i x_j < \vartheta$$

Neurónové siete a vzťah medzi mozgom a mysľou

■ V informatike (menovite v jej časti nazývanej umelá inteligencia) stojíme pred zložitými otázkami, ***čo je ľudská myseľ***, aký je jej vzťah k mozgu, môžeme chápať ľudskú myseľ ako program implementovaný v počítači mozgu, aká je architektúra tohto programu?

■ Žiaľ odpovedať na ne nie je vôbec jednoduché, preto v súčasnosti sú stále predmetom záujmu oblasti filozofie nazývanej „***filozofia mysle***“.

■ Poskytnúť jednoduchý informatický pohľad na vzťah medzi mysľou a mozgom, ktorý je založený na všeobecných predstavách teórie neurónových sietí (nazývanej v kognitívnej vede „***konekcionizmus***“).

■ V rámci *konekcionistického prístupu* vzťah medzi mozgom a mysľou sa rieši pomocou koncepcií modernej neurovedy a neurónových sietí.

V kognitívnej vede je *je mozog/myseľ charakterizovaný ako počítač*

(1) ktorý transformuje symboly pomocou syntaktických pravidiel na iné symboly, pričom

(2) myšlienky sú symbolické reprezentácie implementované pomocou jazyka myslenia, a

(3) mentálne procesy sú postupnosti symbolov (medzi ktorými sú príčinné vzťahy) generované syntaktickými pravidlami.

- Použitie termínu „počítač" obvykle evokuje predstavu sekvenčného počítača von neumannovskej architektúry, kde je možné striktne oddeliť hardware od software; kde na tom istom počítači – hardware môže byť vykonávaných nepreberné množstvo naprosto rozdielnych programov – softvérov.
- Pre tieto počítače existuje striktná dichotómia medzi počítačom a programom – t. j. medzi hardvérom a softvérom.

- Moderný prístup k chápaniu vzťahu medzi mozgom a mysľou je založený na konekcionistickom poňatí tak mozgu, ako aj mysle.

- Základná predstava o mozgu je, že je tvorený z neurónov navzájom poprepájaných pomocou jednosmerných synaptických spojov. Ľudský mozog vykazuje neobyčajnú plasticitu, v priebehu učenia neustále vznikajú (ale taktiež aj zanikajú) synaptické spoje. Architektúra mozgu je určená spojmi medzi neurónmi, ich inhibičným, alebo excitačným charakterom, a taktiež aj ich intenzitou.

- Možno konštatovať, *že schopnosť mozgu vykonávať nielen kognitívne aktivity, ale byť aj pamäťou, je plne zakódovaná do jeho architektúry*.

■ *Myseľ s mozgom tvoria jeden integrálny celok; myseľ je v tomto prístupe možné chápať ako program vykonávaný mozgom, avšak tento program je špecifikovaný architektúrou distribuovanej neurónovej siete reprezentujúcej mozog.*

■ Program v tomto paralelnom počítači je priamo zabudovaný do architektúry neurónovej siete, t. j. ľudský mozog je jednoúčelový paralelný počítač reprezentovaný neurónovou sieťou, ktorý nie je možné preprogramovať bez zmeny jeho architektúry.

■ Na základe týchto neurovedných poznatkov bazálneho charakteru môžeme konštatovať, že počítačová paradigma ľudského mozgu/mysle sa musí formulovať tak, že mozog je paralelný distribuovaný počítač (obsahujúci mnoho miliárd neurónov, elementárnych procesorov, ktoré sú medzi sebou poprepájané do zložitej neurónovej siete).

Záver

(1) Každá Boolova funkcia je reprezentovaná ***neurónovou sieťou***, ktorej architektúra je určená syntantickým stromom funkcie. Táto vlastnosť môže byť zosilonená do tvrdenia

(2) Každá Boolova funkcia je reprezentovaná ***3-vrstvovou neurónovou sieťou***.

(3) Logický neurón klasifikuje len Boolove funkcie, ktoré sú ***lineárne separovateľné***.

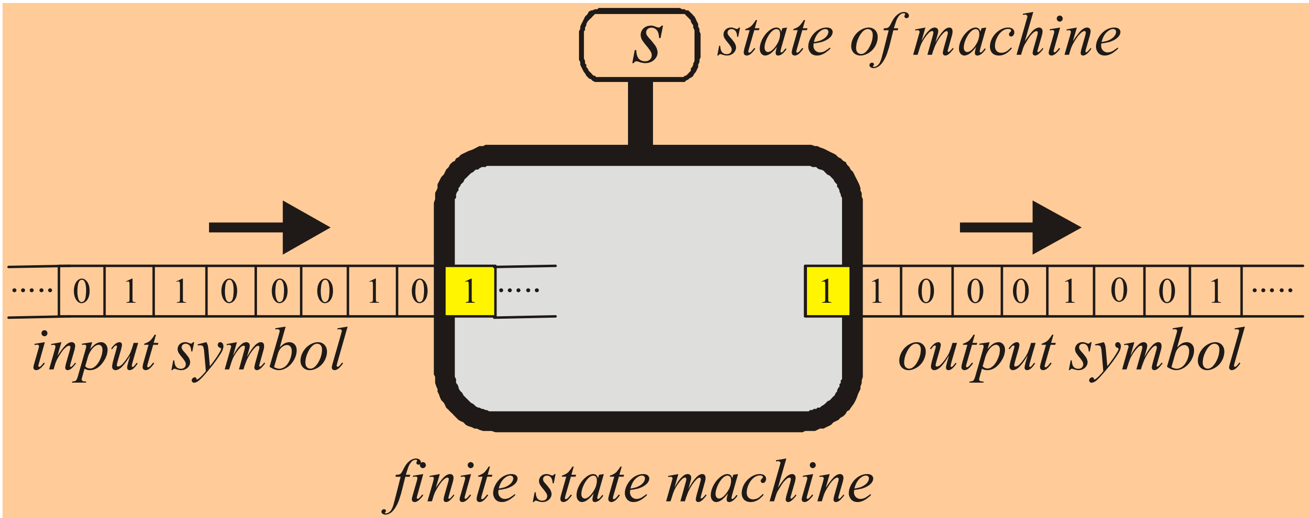
(4) Boolove funkcie, ktoré nie sú lineárne separovateľné, ale sú kvadraticky, kubicky, ... separovateľné, sú reprezentované ***logickým neurónom 2., 3., ... rádu***.

(5) Každá Boolova funkcia je reprezentovaná logickým neurónom vyššieho rádu.

(6) Neurónové siete s logickými neurónmi ***nie sú schopné učenia***, ich architektúra a váhové koeficienty spojov sú fixné.

3. Finite state machine

Marvin Minsky



References

- Kleene, S. C.: Representation of events in nerve nets and finite automata. In C. E. Shannon and J. McCarthy, editors, *Automata Studies*. Princeton University Press, Princeton, 1956, pp. 3-41.
- Minsky, M.: Some universal elements for finite automata. In C. E. Shannon and J. McCarthy, editors, *Automata Studies,* Princeton University Press, Princeton, 1956, pp. 117- 128.
- Minsky, M. L.: *Computation. Finite and Infinite Machines*. Prentice-Hall, Englewood Cliffs, NJ, 1967.

Finite state machine works in discrete time events 1, 2,...,t , t+1,... . It contains two tapes: of input symbols and of output symbols, whereas new states are determined by input symbols and an actual state of machine.

$$state_{t+1} = f\left(state_t, input\ symbol_t\right)$$
$$output\ symbol_{t+1} = g\left(state_t, input\ symbol_t\right)$$

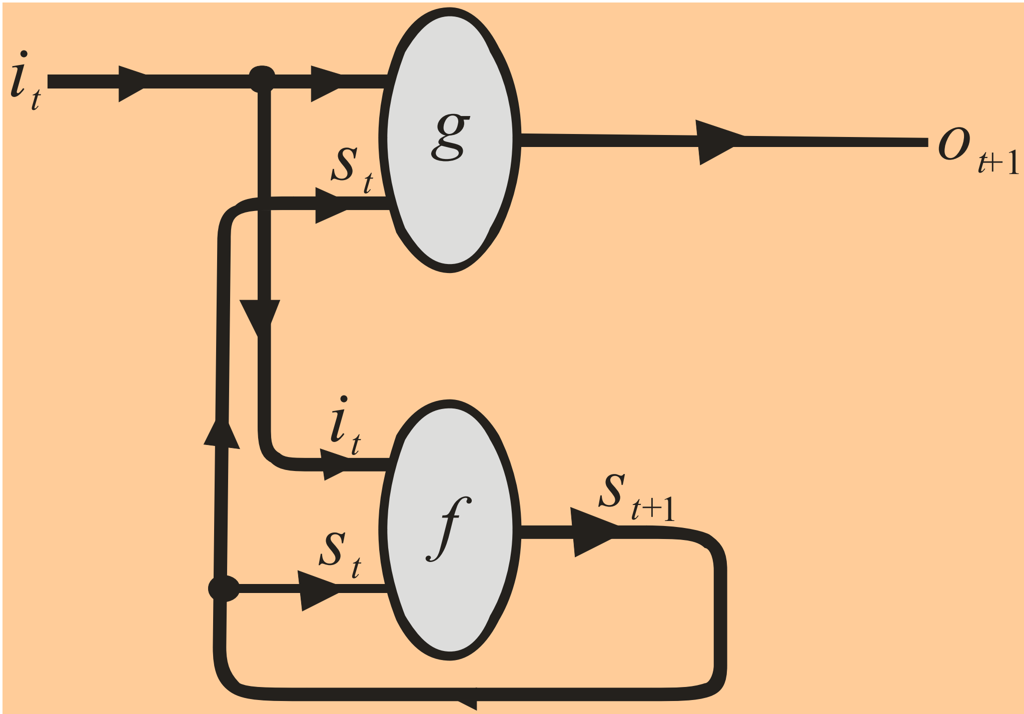
where function f and g specified given finite state machine and are understand as its basic specification:

(1) ***transition function*** f determines forthcoming state from an actual state and an input symbol,
(2) ***output function*** g determines output symbol from an actual state and an input symbol.

Representation of finite state machine as a calculating devise

$$o^{(t+1)} = g\left(s^{(t)}, i^{(t)}\right)$$

$$s^{(t+1)} = f\left(s^{(t)}, i^{(t)}\right)$$



Theorem 1. Any neural network may be represented by equivalent finite-state machine with output.

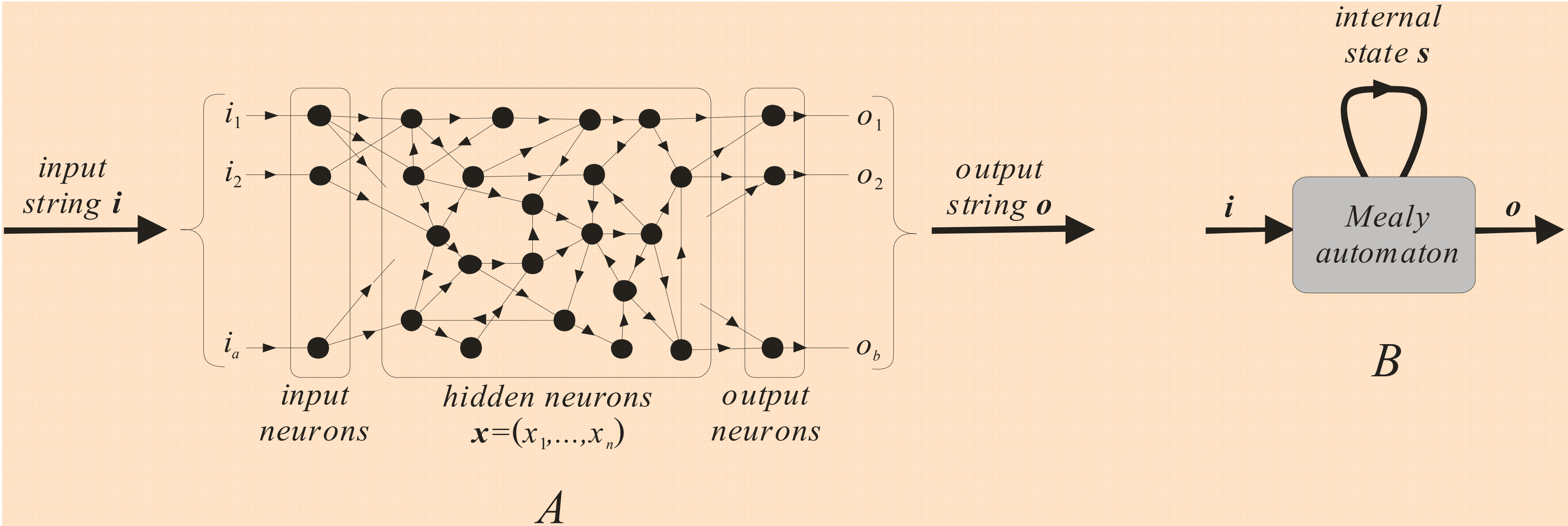
Theorem 2. Any finite-state machine with output (Mealy automaton) may be represented by equivalent recurrent neural network.

Recurrent neural network, which represents a finite-state machine

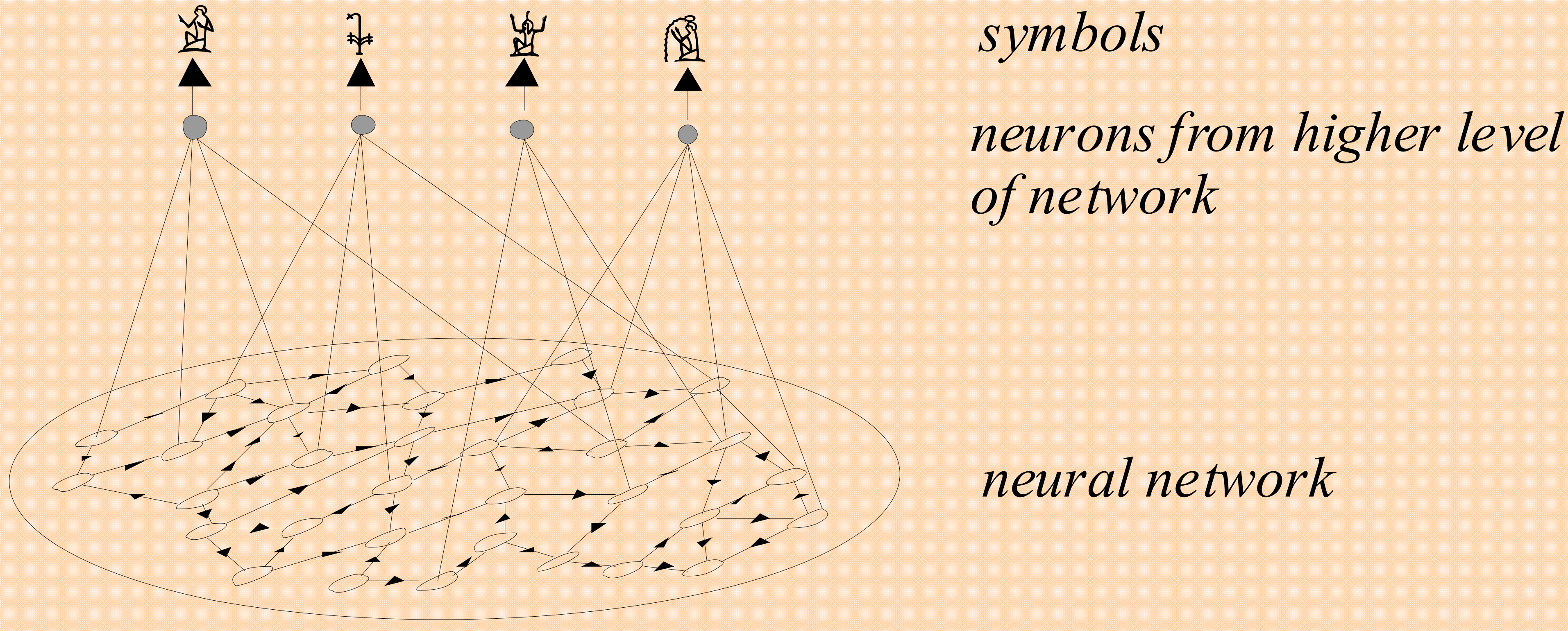
Theorems 1 and 2 make possible to study a classical problem of connectionism (neural networks), a relationship between a ***subsymbolic representation*** (neural, which is represented by patterns composed of neural activities) and a ***symbolic representation*** (which is used by classical AI):

(1) According to the Theorem 1, each subsymbolic neural network can be transformed onto symbolic finite-state machine, whereas symbols may be created by making natural numbers that are assigned to binary vectors of activities.

(2) According to the Theorem 2, each symbolic finite-state machine may be transformed onto an equivalent subsymbolic neural network. Symbols from the representation of finite-state machine are represented by binary vectors.



Above mentioned theory for the transfer from subsymbolic to symbolic representation and conversely may be used for a study of properties of these models and a transfer among them.



Smolensky's idea of hierarchically ordered neural networks, where neurons from the top level of network represent symbols.

Main conclusions from the lecture

- Symbolic vs. subsymbolic (connectionist) representations applied in AI and CogSci ***are not contradictory paradigms*** that are mutually eliminated, rather they are ***complementary dual***.

- Symbols are ***grounded*** through activity of neurons

- Connectionist solution of mind-body problem, a brain may be interpreted as special ***highly-parallel neural computer***, where running program (mind) is wired in via its connectionist structure