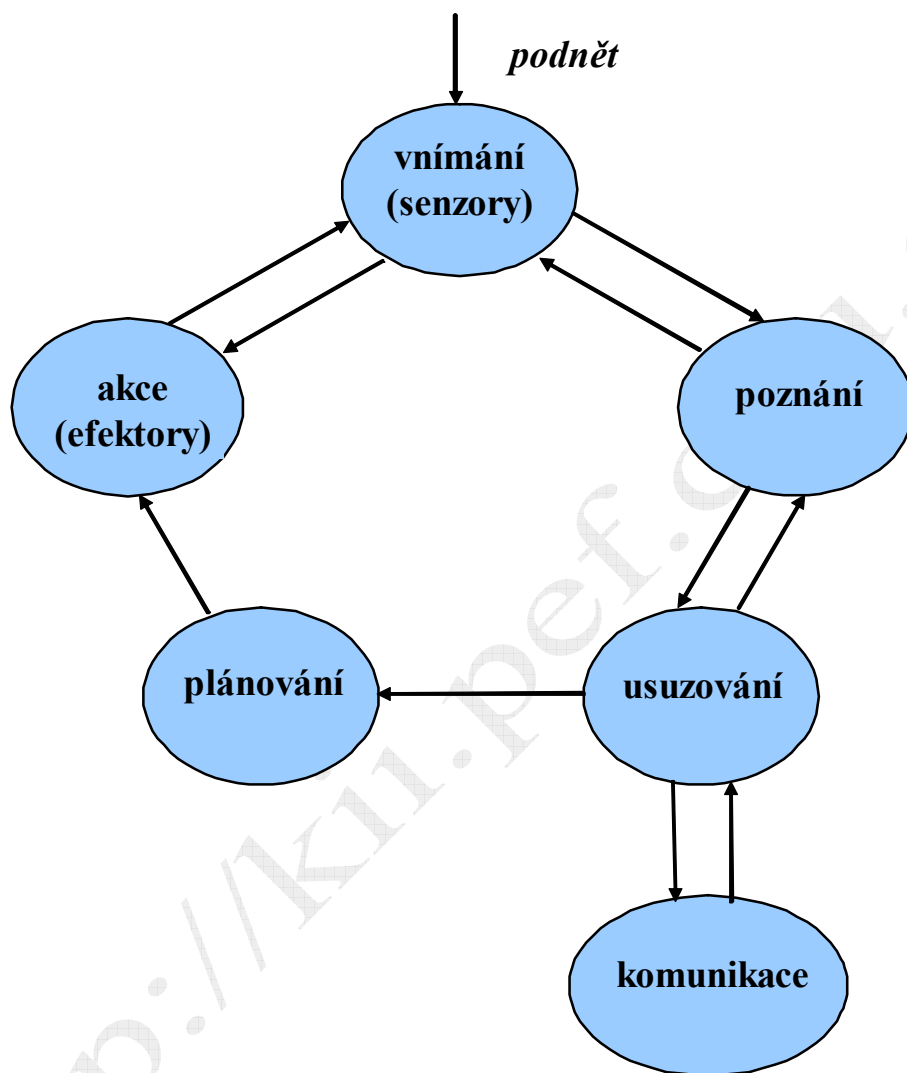


Neuropočítače

Princip inteligentního systému

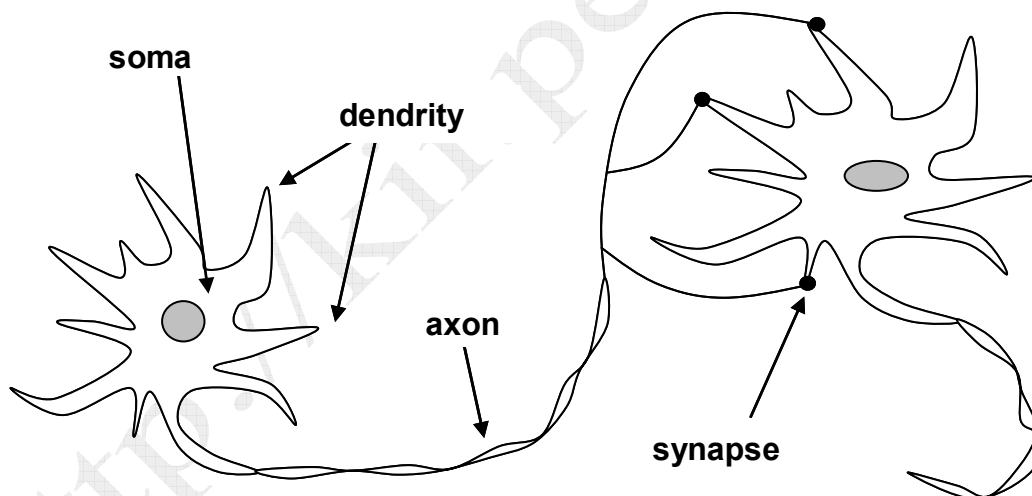


■ Typické vlastnosti **inteligentního systému**:

- schopnost **vnímat podněty** z okolního světa i o svém vlastním stavu – **podráždění**
- **vhodná reakce** na podněty
- schopnost **adaptace** – **učení se**
- **usuzování výsledku** na základě zkušeností – **pamatování si** (kódování) předchozích případů

Informační systém živých organismů

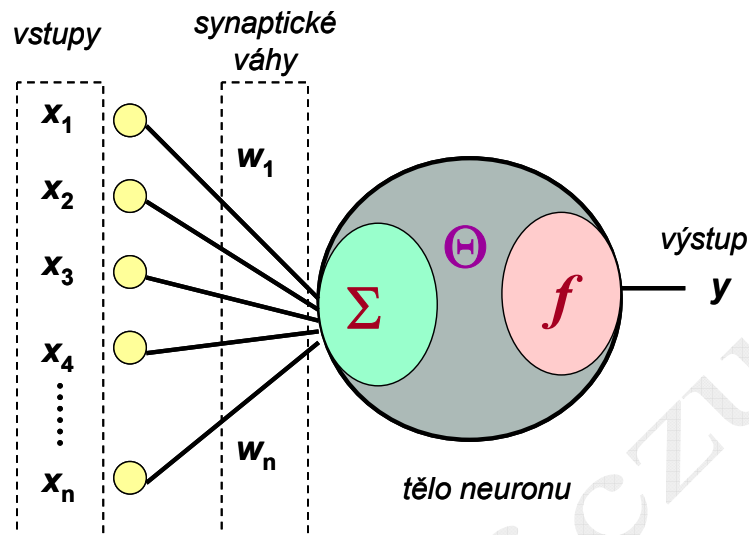
- Předobrazem umělých inteligentních systémů je **nervový systém** a **mozek** člověka
- Základním stavebním prvkem – **neuron** (nervová buňka)
 - většina z nich **přenáší vzruch** uvnitř organismu
 - některé jsou specializovány na přenos informací z vnějšího prostředí – **receptory**
 - některé neurony mají svůj výstup napojen na svalová vlákna, která převedou nervový vzruch na pohyb svalu – **motorické neurony**
- **Biologický neuron má tyto části:**
 - **tělo neuronu (soma)**
 - s buněčným jádrem uvnitř, vyplněné tzv. plazmou obklopenou velice tenkou membránou
 - **dendrity**
 - krátké a tenké, stromčkovitě rozvětvené výběžky, které vedou vzruch směrem k buňce a představují tedy její vstupy
 - **axony**
 - představují funkčně výstup z neuronu
 - jeho délka se u člověka pohybuje v rozmezí několika mikrometrů až jednoho metru
 - axon u každého neuronu jen jeden
 - na jejich koncích pak působí **synapse** – napojení na ostatní neurony



- **Funkce biologického neuronu:**
 - Neuron se chová jako **sumační zařízení** - sčítá jednotlivé signály přivedené přes synapse a tím dochází k jeho podráždění
 - Velký význam zde přitom má tzv. **práh** (prahová hodnota)
 - Je-li podráždění dostatečně silné (nadprahové), dojde ke stavu neuronu se nazývanému **excitace**
 - Excitace, která se lavinovitě šíří nervovým vláknem (axonem) jako vlna směrem k synapsím, se nazývá **akční potenciál** (nervový vzruch)
 - **Synapse** mezi neurony se mění v průběhu činnosti neuronové sítě – mění se jejich průchodnost - **váha**

Matematický model neuronu

■ Matematický model neuronu $y = f(\sum w_i \cdot x_i + \Theta)$, $i = 1 \dots n$



■ Vektor synaptických vah $w_1 \dots w_n$ vyjadřuje uložení zkušeností - schopnost adaptace na základě učení

■ Agregace: $h(t) = \sum w_i \cdot x_i + \Theta$

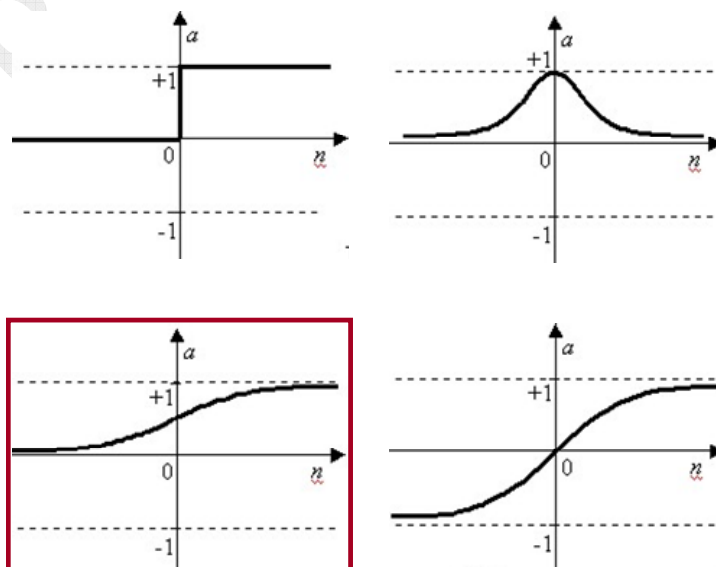
■ jedná se o „vážený“ součet vstupů $x_1 \dots x_n$ a prahové hodnoty Θ

■ Přenosová (aktivační) funkce

■ $y(t) = f(h(t))$

■ převádí na výstup pasivní nebo excitační stav neuronu

■ nejčastěji sigmoidální funkce $y(t) = \frac{1}{1 + e^{-k \cdot h(t)}}$, $y'(t) = k \cdot f(t) \cdot (1 - f(t))$



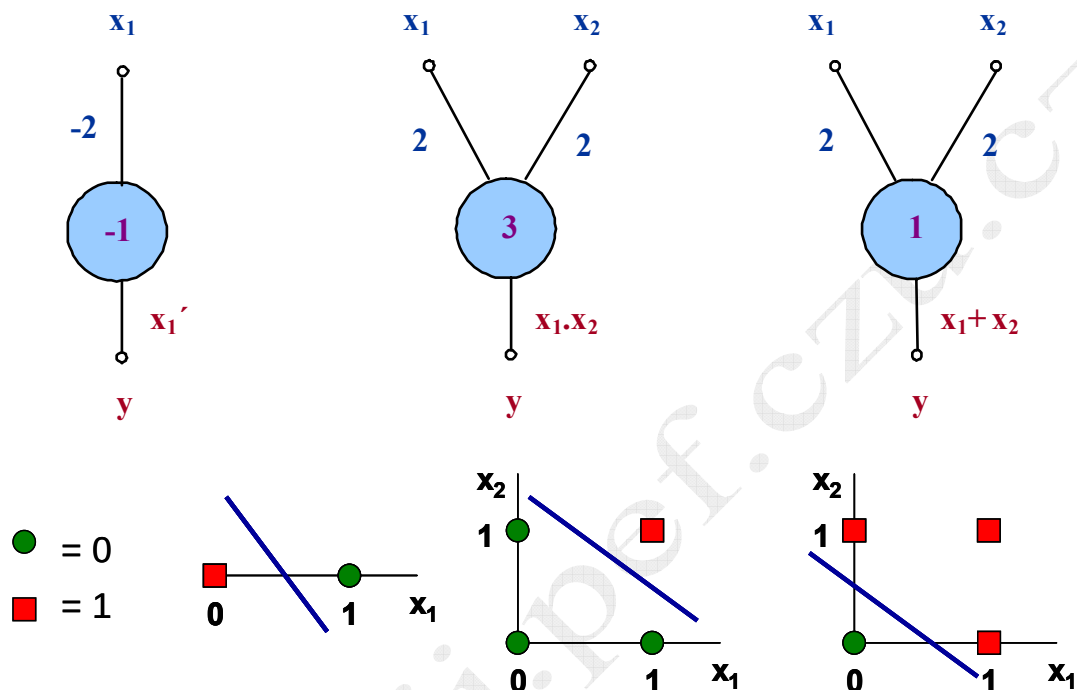
Realizace logických funkcí **NON** (negace), **AND** (logický součin) a **OR** (logický součet) neuronem:

■ přenosová funkce: $f(x) \sim \sigma(x)$

$w_i \cdot x_i > \Theta \rightarrow y = 1$

$w_i \cdot x_i < \Theta \rightarrow y = 0$

■ $x_i \in \{0,1\}$

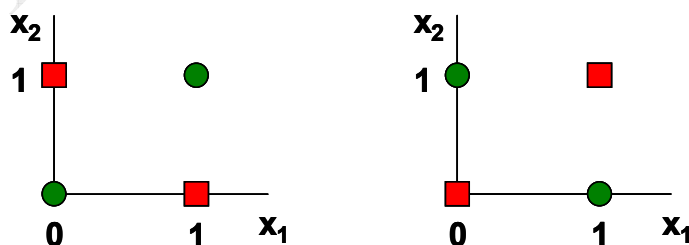


■ Výstupy logických funkcí **NON**, **AND** a **OR** tvoří disjunktivní kategorie oddělitelné přímkou – **lineárně separabilní**

■ **Jedním neuronem lze realizovat každou funkci, pokud je lineárně separabilní**

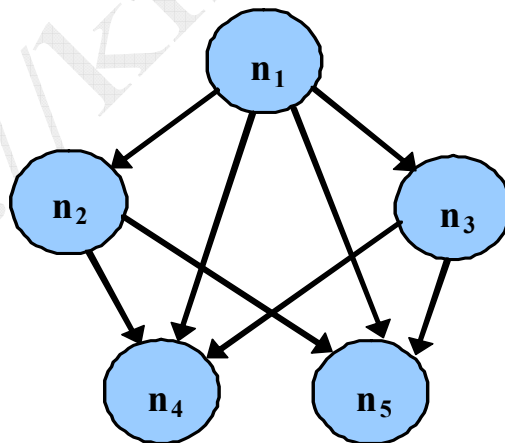
■ Realizace logických funkcí **XOR** (výlučné nebo, nestejnost) a **EQ** (stejnost)

■ jedná se o **lineárně neseperabilní** funkce $\rightarrow$ nelze realizovat jedním neuronem



Neuronová síť

- **Neuronová síť** je paralelní systém výkonných prvků modelující biologické neurony, účelně uspořádaný tak, aby byl schopen požadovaného zpracování informací
 - provádí transformační funkci $Y = T(X)$
- Je-li **reprezentována orientovaným grafem**, pak:
 - **uzly** grafu jsou **neurony** (modely)
 - **hrany** jsou spoje – **jednosměrné přenosové kanály**
 - každý spoj má svoji **váhu**
 - každý uzel **libovolný počet vstupních spojů**
 - každý uzel **libovolný počet výstupních spojů**
 - ve všech výstupech je stejný signál
 - každý uzel má svoji lokální paměť – **práh**
 - každý uzel má svoji **přenosovou (aktivační) funkci**
- Váhami ohodnocený graf se nazývá **konfigurace sítě**
- Z hlediska propojení neuronů v síti rozlišujeme dva základní **typy topologie**
 - **Cyklická (rekurentní) síť**
 - V případě cyklické sítě existuje v grafu topologie uzavřená smyčka, tj. existuje skupina neuronů zapojená do kruhu
 - **Acyklická (dopředná) síť**
 - Acyklická síť je taková síť, jejíž graf neobsahuje žádnou smyčku a lze ji tedy uspořádat do vrstev tak, že spoje mezi neurony vedou jen z vyšších vrstev do nižších
 - Speciálním případem acyklické struktury je tzv. **vícevrstvá neuronová síť**



Životní cyklus neuronové sítě

- V **životním cyklu každé neuronové sítě** můžeme sledovat tři etapy:
 - **organizační etapa** – stanovení typu struktury neuronové sítě a její velikosti
 - na základě znalostí navrhovatele nebo v rámci procesu učení
 - **adaptační etapa** (trénování, učení) – adaptace parametrů sítě (**váhy, prahy**) s ohledem na cíl, který má síť plnit
 - **aktivační etapa** (vybavování, odezva) – po předložení určitého vstupu (**podnětu**) je dosaženo ustáleného stavu na výstupu (**odezvě**)
- Podle míry autonomie učícího procesu můžeme rozlišovat tři **hlavní typy učení**
 - **učení s učitelem**
 - Během učení musí být síti předkládán pro každý vstup také odpovídající požadovaný výstup, který je porovnán s výstupem skutečným a na základě vyhodnocené chyby jsou prováděny korekce parametrů sítě
 - **hodnocení učení**
 - Místo předkládání požadovaného výstupu je prováděno pouze hodnocení, do jaké míry je skutečný výstup přiměřený, přičemž hodnocení není třeba provádět pro každý vstup, ale jen v určitých časových intervalech
 - **učení bez učitele**
 - Síť s takovým typem učení vyžadují pouze předkládání vstupů a učí se dávat výstupy určitým způsobem odpovídající vstupům. Tyto sítě si vytváření svůj vlastní názor a jejich funkce může být v rozporu s našimi původními požadavky

Model vícevrstvé neuronové sítě

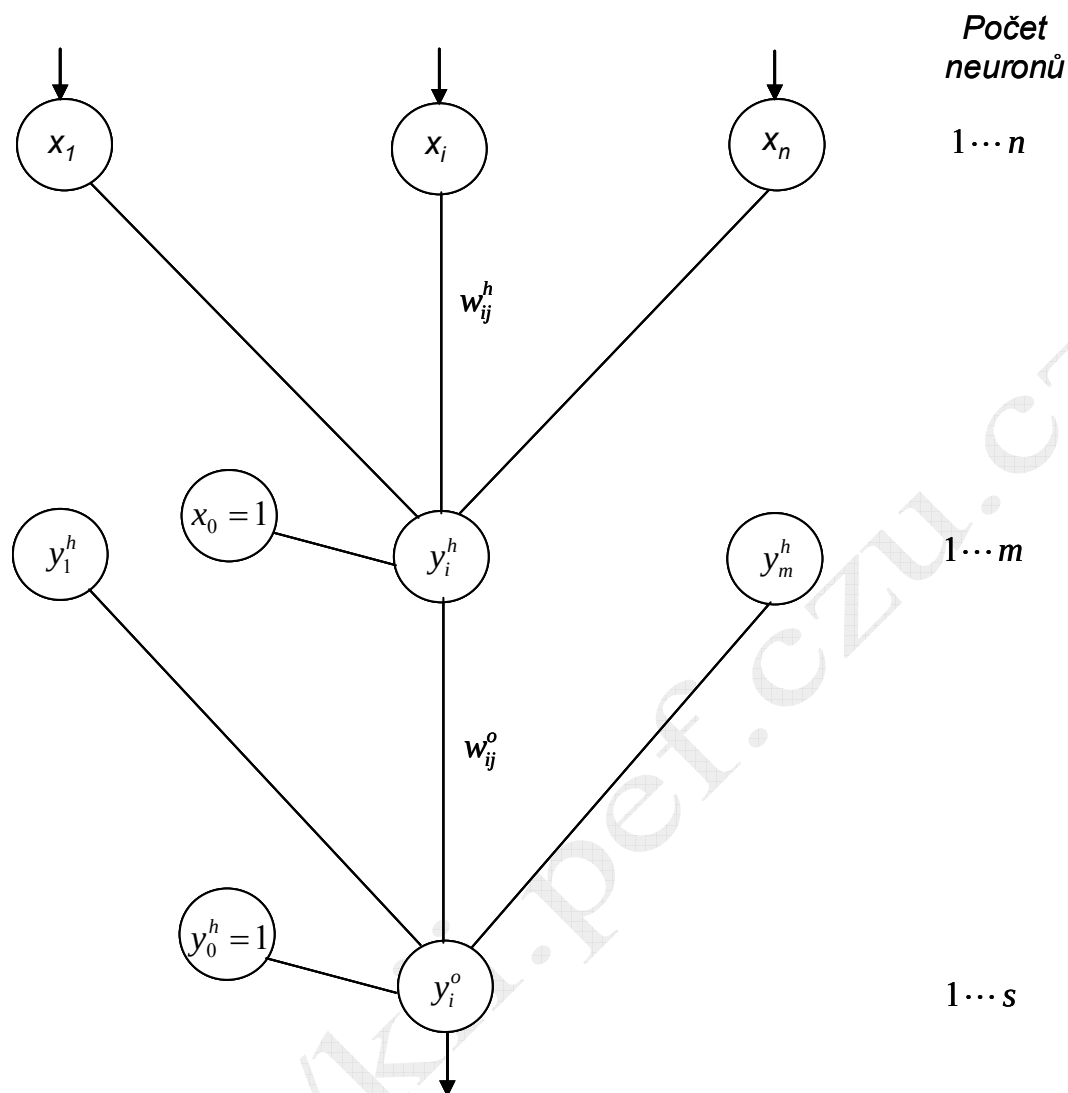
- **Vícevrstvá neuronová síť** se skládá z těchto vrstev neuronů:
 - **vstupní vrstva** (*input layer*)
 - **skrytá vrstva** (*hidden layer*) - jedna nebo více
 - **výstupní vrstva** (*output layer*)

Adaptace vrstvených neuronových sítí

- Metoda adaptace vrstvených neuronových sítí, která patří mezi metody učení s učitelem, je známa jako tzv. **algoritmus zpětného šíření chyby** (*back-propagation*)
 - Jedná se v podstatě o **gradientní metodu minimalizující součet čtverců chyby výstupních hodnot** dané sítě
 - Chceme tedy minimalizovat chybovou funkci, které se také říká **energie neuronové sítě**
 - Neuronová síť považována za naučenou, je-li skutečný výstup byl co nejvíce shodný s výstupem požadovaným
- Předpokládejme, že síti budeme postupně předkládat **učební soubor D** ve tvaru:

$$D = \{(x_1, x_2, \dots, x_n), (d_1, d_2, \dots, d_s)\}$$

- kde $x_i, i = 1, \dots, n$ jsou vstupy
- $d_j, j = 1, \dots, s$ jsou požadované výstupy



■ **Energetická funkce** je definované následovně:

$$E = \frac{1}{2} \sum_{i=1}^s (y_i^o - d_i)^2$$

■ kde y_i^o , $i = 1, \dots, s$ jsou výstupy sítě

■ Naším cílem je tuto hodnotu snížit modifikací váhových vektorů

■ **Energetickou funkci** si lze představit jako zakřivenou plochu v prostoru

■ Okamžitá hodnota synaptických vah je pak bodem na této ploše

■ Z tohoto bodu se snažíme postupně dosáhnout **minima energie**

■ Chceme-li dojít do energetického minima, musíme jít podle největšího spádu
– **gradientu energetického prostoru**:

$$\text{grad } E = \frac{\partial E}{\partial w_{ij}}$$

- Z gradientu energetického prostoru můžeme vypočítat **vektor přírůstku vah**:

$$\Delta \bar{w} = -\eta \text{ grad} E$$

- Váhy (ve všech vrstvách sítě) upravujeme vždy η **násobkem záporné hodnoty gradientu** - změna znaménka je důležitá, neboť **gradient je orientován směrem k vrcholu**
- váhy sítě upravují po každém učebním kroku, kdy jsou předkládány prvky učebního souboru v náhodném pořadí
- Při praktických aplikacích algoritmu zpětného šíření jsou často problémy s jeho **konvergencí**
 - Metoda je velmi citlivá na nastavení optimalizačního parametru η , tzv. **kroku učení**
 - Nebezpečí **uvážnutí** algoritmu pouze **v lokálním** a ne absolutním **minimu**

$$\Delta \bar{w}(t) = -\eta \cdot \text{grad} E + \alpha \cdot \Delta \bar{w}(t-1)$$

- Parametr α přidává do postupu energetickým prostorem tzv. **setrvačnost** - takto je možné lokální minimum překonat
- Významným faktorem, který výrazně ovlivňuje výsledek učení je **počáteční nastavení vah**

Aplikace neuronových sítí

- Teoretické výsledky ukázaly, že **k realizaci libovolné reálné funkce** stačí (vrstvená) **neuronová síť s jednou skrytou vrstvou**
- **Praktické použití neuronových sítí** závisí na oblasti řešených úloh
 - **Vhodné oblasti použití**
 - všechny úlohy týkající se zpracování neúplných, nepřesných, rozporných a neurčitých informací, rozpoznávání složitých signálů a obrazů a též úlohy optimalizačního charakteru za obtížných, často časově značně proměnlivých podmínek
 - Jedná se především o řešení problémů, kde není znám jednoznačný algoritmus řešení, ale existuje dosti velká množina příkladů, jejichž řešení známe
 - **Nevhodné oblasti použití**
 - neuronové sítě nejsou vhodné pro přesné výpočty, pro zpracování numerických hodnot, zejména pak úloh rozsáhlejších
- Při aplikaci neuronových sítí je třeba **věnovat pozornost**
 - **formulaci problému**, jeho rozsahu a požadované přesnosti a rychlosti řešení
 - **času**, který je k dispozici **pro učení** příslušné sítě
 - zda je ji možno naučit jednou pro vždy či alespoň pro dostatečně velkou skupinu zadání nebo zda je její učení nutno provádět při každém zadání znovu
 - požadavkům na **rychlost odezvy** již naučené sítě
 - **výběru učebních a testovacích vzorů**

- **Typické aplikace neuronových sítí** je možno rozdělit do několika větších skupin
 - Analýza a klasifikace jednorozměrných signálů
 - Predikce časových řad
 - Rozpoznávání obrazů
 - Filtrace šumu
 - Řízení složitých zařízení
 - Syntéza a rozpoznávání řeči
 - Znalostní systémy
 - Optimalizace

Realizace neuronových sítí

- **Umělé neuronové sítě jsou charakterizovány**
 - **svoji strukturou** – počtem a propojením aktivních prvků (uzlů – neuronů)
 - **hodnotami parametrů**
 - **synaptické váhy**
 - lze reprezentovat synaptickými průchodnostmi nezávislými na vnějším prostředí a čase, měnící se pouze v procesu učení
 - vyjadřujeme obvykle **reálnými čísly** (váhami)
 - **přenosové funkce**
 - lze reprezentovat jednoduchými matematickými funkcemi
 - parametry těchto funkcí vyjadřujeme **opět reálnými čísly**
 - **prahové funkce**
 - obvykle zjednodušovány na **reálná čísla** (práh)
- Umělé neuronové sítě jsou vytvářeny ze **dvou hlavních důvodů**
 - jako nástroj pro modelování biologických neuronových sítí
 - jako nástroj pro řešení různých praktických problémů
- **Prostředky pro realizaci** lze rozdělit do dvou hlavních kategorií
 - **prostředky simulací na počítačích**
 - **prostředky fyzické realizace**

Prostředky simulací na počítačích

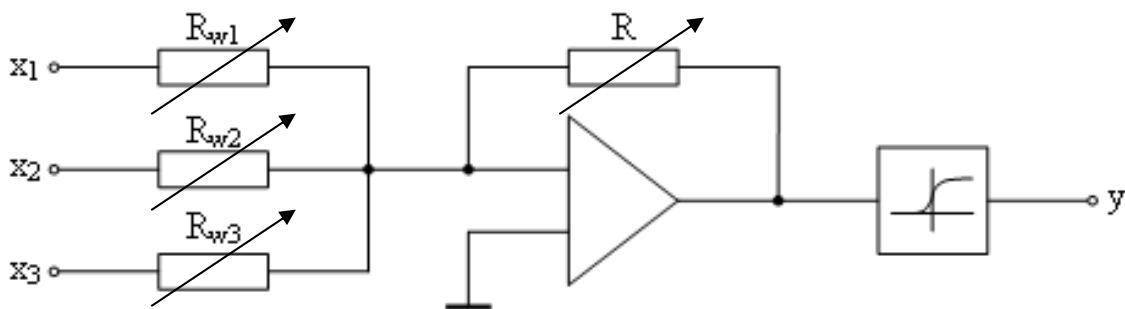
- Při simulaci umělých neuronových sítí na počítačích je třeba brát v úvahu tato **kritéria**
 - **jak velkou a složitou síť můžeme simulovat**
 - **jak rychle je možno získat odezvu**
 - **jak dlouho bude trvat učení sítě** - proces učení je daleko pomalejší než proces vybavování
- Na běžných počítačích (PC) je největším omezením jejich **sekvenčnost** – je vyžadováno provádění velkého množství numerických operací
 - lze použít paralelní systémy (superpočítače) – obvykle ale nákladné řešení
- Pro simulaci neuronových sítí byla vyvinuta celá řada **programů**
 - NESP (AV ČR), Software Environment for Neural Networks, Neural Network Toolbox, Stuttgart Neural Network Simulator, ...

Prostředky fyzické realizace

- **Fyzická realizace** vede v naprosté většině k integraci neuronové sítě na jednom čipu – **neuročipu**
- Podle režimu, ve kterém neuročipy pracují rozlišujeme
 - **analogové a digitální**
- **Požadavky na vlastnosti neuročipů**
 - schopnost pracovat s **libovolným paradigmatickým neuronové sítě**
 - rozsah **řádově stovky až tisíce** aktivních prvků
 - možnost rychlého a co možná nejméně omezeného **propojení čipů navzájem**
 - možnost pracovat s **digitálními i analogovými signály**
- Mezi **obecné problémy fyzické realizace** neuronových sítí patří
 - **adaptivní nastavování synaptických vah**
 - nejčastěji pomocí proměnného nastavení vodivosti rezistorů ve vstupních spojích
 - problém přesnosti a časové stálosti
 - **omezený možný počet vývodů z čipu**
 - každý čip je nutné zapouzdřit s nutně omezeným počtem vývodů
 - počet vývodů omezuje složitost sítě
 - **omezený počet neuronů na jednom čipu**
 - vede k potřebě je navzájem skládat do mozaiky tvořící vlastní síť
- Neuronové sítě realizované v **neuročipech** mohou být
 - **analogového charakteru**
 - poměrně levné stavební prvky
 - velká rychlost a paralelní zpracování signálu
 - menší stupeň univerzality – nelze např. použít složitější přenosové funkce
 - problém přesného a stabilního nastavení vah
 - **digitálního charakteru**
 - složitější speciální obvody
 - binární reprezentace váhových koeficientů
 - cenově náročnější

Analogová implementace

- **Analogová implementace** neuronové sítě znamená sestavení **každého neuronu z analogových prvků**
 - **Neuron = sumátor** s váženými vstupy
 - **Rezistory** = analogové násobičky $x_i \cdot w_i$
 - proměnné hodnoty vah (w_i)
 - **vysoká rychlost a paralelní zpracování signálů vs nízká přesnost**

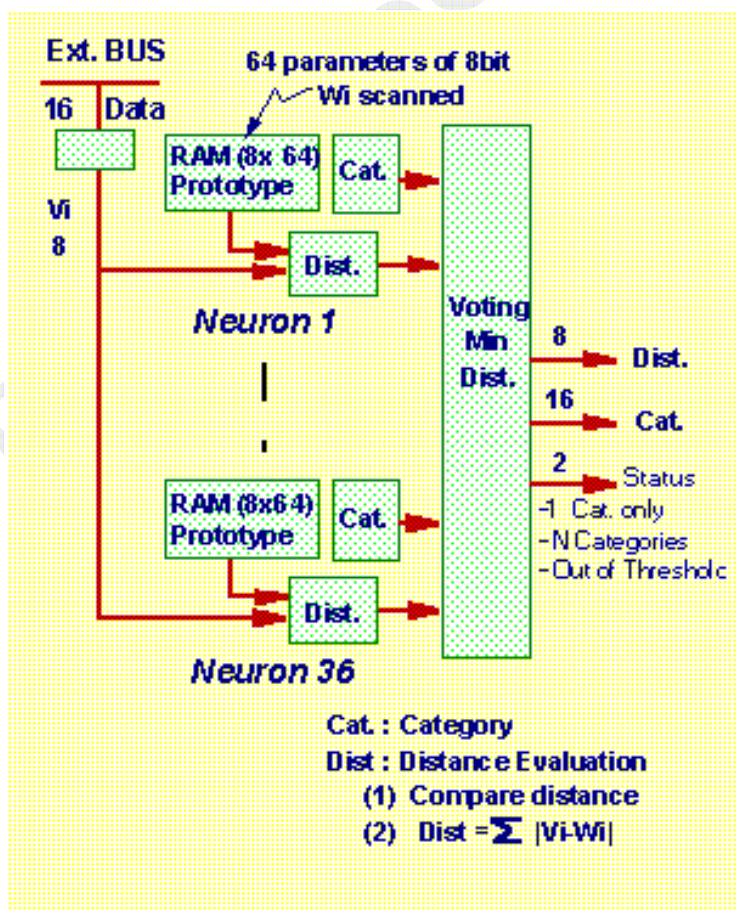


Digitální implementace

■ **Digitální implementace** představuje

■ **číslicové neuročipy**

- specializovaný integrovaný obvod obsahující stavební prvky pro implementaci digitálních neuronových sítí
- cenově i konstrukčně vhodné
- **MD 1220** (8 neuronů), **ZISC 36** (36 neuronů)
- interní nebo externí učící algoritmus
- omezení pouze na **binární vstupy a výstupy**

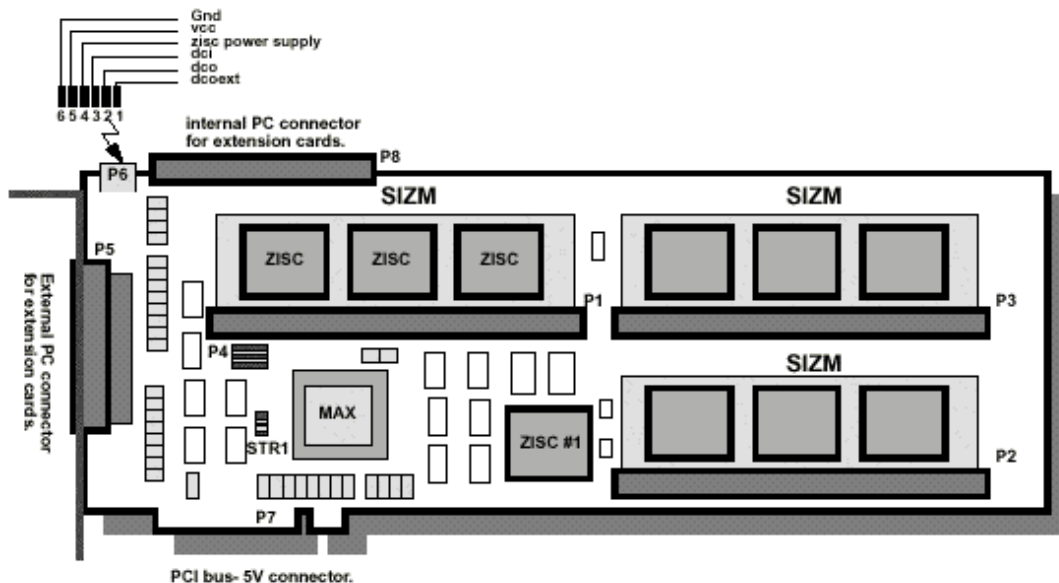


■ neuronový akcelerátor

- specializovaný integrovaný obvod pro simulaci neuronové sítě
- vybaven signálovým nebo vektorovým procesorem → rychlé provádění vektorových operací
- obvykle ve formě přídavné karty (PCI)

■ ZISC akcelerátor

- složen z neuročipů ZISC36 – až 684 neuronů (SIZM Single In-line ZISC Module)
- lze navzájem propojit i více karet (PCI)
- ovládání přes dynamickou knihovnu (DLL)



Příklad neuronového systému

■ Struktura - 40-100 miliard neuronů

- každý neuron 10-100 tisíc spojení, celkem 10^{15-16}

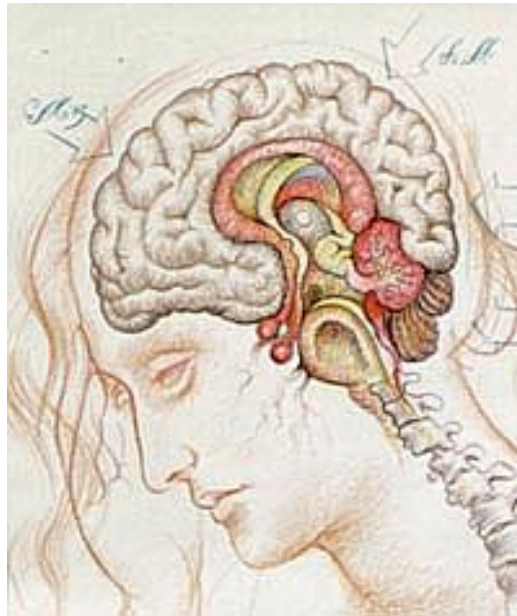
■ Topologie - vrstevnatá oscilační pyramidální síť

- 6 vrstev
- informace se zpravidla šíří dopředu - postupné snižování počtu signálů ve směru šíření
- paralelní (lavinovitý) přenos informací, přenos mezi dvěma neurony 20-40 ms
- složité zpětné vazby – i přes několik vrstev

■ Paměť

- **dynamická** – různé, proměnlivé stavy
- cirkulace informace po uzavřených okruzích neuronů – **paměťové stopy** → změna vah i vlastností neuronů
- velmi vysoká kapacita – až neomezená

- **Adaptace** - schopnost učení se
 - **neasociativní** (jednoduché stimuly bez souvislostí) i **asociativní** (vztahy mezi stimuly)
 - **s učitelem** (pomocí vhodných vzorů), **bez učitele** (bez vzorů) i **hodnocené učení** (bez vzorů ale hodnocené výsledky učení)
- **Biologická „implementace“** – lidský mozek



- Současné umělé neuronové sítě si nemohou činit nárok na bližší napodobení funkcí lidského mozku jako je vnímání, vědomí, emoce, ...
 - použitelné pouze pro dílčí a daleko jednodušší funkce