1 História neurónových sietí a inšpirácie z neurobiológie

Neurónové siete (ktoré sa v kognitívnej vede[1] nazývajú „*konekcionizmus*“) v súčasnosti patria medzi významnú časť počítačovo orientovanej umelej inteligencie, kde zaujali postavenie univerzálneho[2] matematicko-informatického prístupu k štúdiu a modelovaniu procesov učenia, adaptácie umelých kognitívnych systémov založených na metafore ľudského mozgu. Okrem umelej inteligencie neurónové siete majú nezastupiteľné uplatnenie aj v kognitívnej vede, lingvistike, neurovede (neuroscience), riadení procesov, prírodných a spoločenských vedách, kde sa pomocou nich modelujú nielen procesy učenie a adaptácie, ale aj široké spektrum rôznych problémov klasifikácie objektov a taktiež problémov riadenia zložitých priemyselných systémov. V tejto súvislosti musíme upozorniť, že najväčší a principiálny význam majú neurónové siete v neurovede a v kognitívnej vede, kde patria medzi základné teoretické metódy pre interpretáciu rôznych aktivít nášho mozgu. V týchto dvoch oblastiach vznikli základné konekcionistické teoretické prístupy (neurónové siete) a bola preukázaná ich vhodnosť a efektívnosť pre štúdium a modelovanie najrozličnejších aktivít a aspektov ľudského mozgu. Konekcionizmus reprezentuje dôležitý pojmový a argumentačný aparát, ktorý umožňuje interpretovať a vysvetľovať kognitívne aktivity ľudského mozgu spôsobom, ktorý je v súlade s našimi predstavami o štruktúre a fyziológii mozgu. O význame a postavení konekcionizmu v systéme kognitívnych a informatických vied diskutuje Ivan Havel v jeho článku venovanom filozofickým problémom myslenia [7].

V rámci dejín konekcionizmu možno rozlíšiť tieto etapy teórie:

1. etapa - *obdobie klasického konekcionizmu* rozvíjaného najmä v psychológii; vyvrcholením tejto etapy je práca McCullocha a Pittsa (1943) o logických neurónoch [19].

2. etapa - *obdobie počítačového konekcionizmu*, v ktorom vznikla teória umelých neurónových sietí zásluhou prác Rosenblatta [23,24] o perceptróne a Rumelharta a spol. [26,27] o učení sa neurónových sietí obsahujúcich skryté neuróny; ktoré umožnili biologicky paluzibilné počítačové modelovanie procesov učenia sa. Je potrebné poznamenať, že konekcionistické modely sú výpočtovo veľmi náročné, bez použitia počítačov nie je možné aplikovať ich na modelovanie kognitívnych procesov prebiehajúcich v ľudskom mozgu.

[1] Kognitívna veda je interdisciplinárneho charakteru, ktorá leží na rozhraní filozofie, neurovedy, psychológie, lingvistiky a umelej inteligencie. Hlavným cieľom kognitívnej vedy je vysvetliť kognitívne procesy ľudskej mysli pomocou teoretických predstáv, ktoré sú plauzibílne so súčasnými predstavami neurovedy, psychológie a umelej inteligencie.

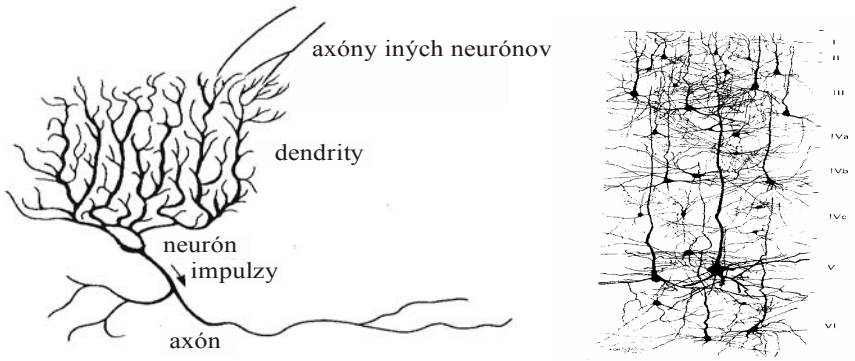
[2] Pod "univerzálnosťou" konekcionistického prístupu sa rozumie skutočnosť, že umelé neurónové siete majú nasledujúce tri všeobecné vlastnosti: (1) Neurónová sieť s dopredným šírením signálu a s aspoň jednou vrstvou skrytých neurónov je *univerzálny aproximátor* [11]. To znamená, že ľubovoľná kognitívna aktivita, ktorá sa dá vyjadriť pomocou tabuľky (tréningovej množiny) obsahujúcej dvojice vstup/požadovaný výstup, sa dá vypočítať pomocou umelej neurónovej siete. (2) Rekurentná neurónová sieť je schopná simulovať *deterministický konečný automat* [29]. Táto vlastnosť rekurentných neurónových sietí znamená, že sú schopné klasifikovať konečné reťazce znakov (t.j. konekcionizmus má svoje interné prostriedky na priamu manipuláciu so symbolickou informáciou). (3) Idealizované neurónové siete sú buď *turingovsky ekvivalentné alebo majú "super-turingovskú" výpočtovú silu* [28]. Žiaľ, táto posledná vlastnosť nemá priamy dopad na aplikáie konekcionizmus v umelej inteligencii a v kognitívnej vede.

Možno konštatovať, že umelá inteligencia, kognitívna veda a neuroveda získali v moderných neurónových sieťach mocný a efektívny simulačný nástroj na modelovanie kognitívnych procesov. Žiaľ, výpočtová náročnosť týchto modelov je taká veľká, že ich použitie na modelovanie je možné len prostredníctvom modernej výpočtovej techniky.

1.1 Základné princípy neurónových sietí - inšpirácie z neurobiológie

Konekcionizmus v umelej inteligencii (alebo v kognitívnej vede) sa chápe ako spôsob paralelného spracovania informácie. Na rozdiel od klasického - *symbolického prístupu*, kde sa sériovo pracuje so symbolmi pomocou hierarchicky usporiadaných logických pravidiel, v konekcionistickom prístupe sa uplatňuje paralelné spracovanie informácie pomocou jednoduchých výpočtov realizovaných neurónmi. V konekcionistickom prístupe je informácia reprezentovaná "z pohľadu" jednotlivých neurónov v sieti jednoduchým sledom impulzov, kde je dôležité, ktoré neuróny v rámci štruktúry siete sú aktívne. Konekcionistické modely sú založené na metafore ľudského mozgu, interpretujú a modelujú kognitívne vlastnosti mozgu pomocou teoretických predstáv, ktoré majú svoj pôvod v neurovede. V konekcionizme sa vychádza zo základného postulátu neurovedy, že základným stavebným kameňom ľudského mozgu je neurón, ktorý má tieto základné vlastnosti [2,16,17]:

(1) neurón *prijíma signály* z okolia od ostatných neurónov,
(2) neurón *spracováva* (*integruje*) prijaté signály,
(3) neurón *posiela spracované signály iným neurónom zo svojho okolia*.



Obrázok 1.1. Ľavý obrázok znázorňuje typickú neurónovú bunku, ktorá obsahuje rozsiahly dendritický systém a dlhý vetviaci sa axón. Prostredníctvom dendritického systému do neurónu vstupujú signály z iných neurónov a prostredníctvom axónu z neurónu vystupuje signál charakterizujúci stav neurónu. Pravý obrázok znázorňuje vzájomné prepojenie neurónov pomocou spojov medzi dendritmi a axónmi.

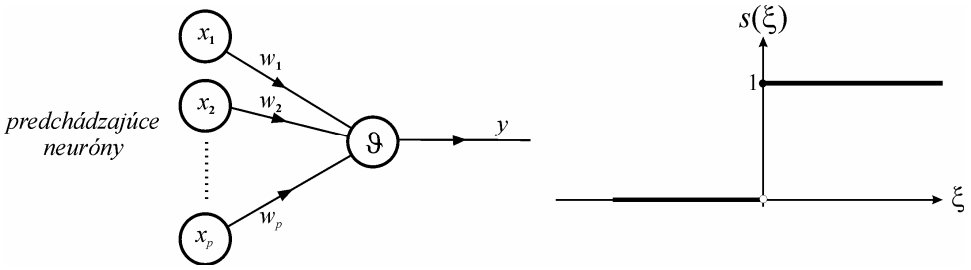
Toto sú tri základné funkcie biologického neurónu, ktoré tvoria kauzálny reťazec *vstup - transformácia - výstup* (pozri ľavú časť obr. 1.1). Neurónová sieť (mozog) je tvorená z neurónov, ktoré sú poprepájané pomocou rozvetvených vstupných a výstupných orgánov jednotlivých neurónov (pozri pravý obr. 1.1). Táto biologická predstava o práci neurónu našla svoj odraz aj v slávnej dvojdielnej publikácii od Rumelharta a kol. "*Parallel Distributed Processing*" [25], ktorá sa pokladá za primárny zdroj moderného konekcionizmu (teórie umelých neurónových sietí, ako sa tento odbor často označuje v informatike). Od konekcionistických modelov sa požadujú tieto vlastnosti:

(1) model je zložený z procesných jednotiek - (umelých) neurónov,
(2) neuróny majú stavy aktivácie,
(3) model obsahuje sieť spojov medzi neurónmi,
(4) neuróny sú charakterizované aktivačnými pravidlami, ktoré opisujú výstupnú aktiváciu daného neurónu pomocou jeho vstupných aktivít,
(5) model obsahuje pravidlo učenia sa, pomocou ktorého sa modifikujú váhy spojov v sieti neurónov tak, aby výstupné aktivity boli blízke požadovaným,
(6) model je aktívny v nejakom prostredí, z ktorého prijíma vstupné signály.

Podľa týchto požiadaviek (pozri vlastnosť 4) konekcionistický model obsahuje tzv. *aktivačnú funkciu*, pomocou ktorej sa transformujú vstupné signály na výstupný signál. Jednoduchý model tejto aktivačnej funkcie má tvar krokovej funkcie (pozri obr. 1.2)

$$y = s(\xi) = s\left(\sum_{i=1}^{p} w_i x_i + \vartheta\right) \tag{1.1a}$$

$$s(\xi) = \begin{cases} 1 & (\xi \geq 0) \\ 0 & (\xi < 0) \end{cases} \tag{1.1b}$$



Obrázok 1.2. Ľavý obrázok znázorňuje prácu neurónu, ktorý obsahuje spoje s *p* inými neurónmi. Neurón je buď v stave reprezentovanom 0, potom nevysiela signály do okolia pomocou axónu, alebo je v stave 1, potom pomocou axónu vysiela signály do okolia. Ak váhovaná suma vstupných aktivít je väčšia ako prah -ϑ, tak stav neurónu je *y*=1, v opačnom prípade, ak váhovaná suma je menšia ako prah -ϑ, tak neurón je v stave *y*=0. Pravý obrázok je graf priebehu krokovej funkcie (1.1b).

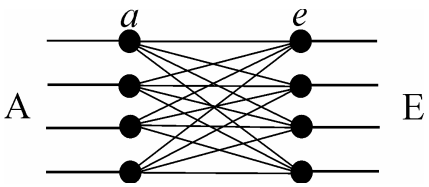
Niekoľko záverečných poznámok o význame vyššie uvedených základných princípov konekcionizmu pre umelú inteligenciu, o všeobecných dôsledkoch z nich bezprostredne vyplývajúcich. Neuróny a ich spoje sú *extrémne jednoduché výpočtové zariadenia*, ktoré sú schopné spracovávať resp. prenášať len sekvencie jednoduchých signálov - impulzov. Predstava o tom, že už na úrovni neurónov je spracovávaná symbolická informácia (t.j. štruktúrovaná inak, ako do jednoduchej sekvencie impulzov) je *a priori* chybná. Preto konekcionizmus v umelej inteligencii stojí pred určitým paradoxom, ako pomocou jednoduchých "subsymbolických" výpočtových jednotiek je možné vysvetliť a interpretovať vlastnosť ľudského mozgu ako celku, ktorý evidentne je schopný nielen manipulovať so symbolickou informáciou, ale je schopný ju aj ukladať a aj vyberať z pamäti. Vzťah medzi symbolizmom a konekcionizmom v umelej inteligencii a v kognitívnej vede nie je jednoduchý, vo všeobecnosti možno konštatovať, že sa jedná o dve rôzne hierarchické úrovne nazerania na spracovanie informácie v neurónových sieťach. Prvý pohľad je mikroskopický – subsymbolický, ktorý používa pojmy a koncepcie na úrovni neurónov a ich spojov. Druhý alternatívny pohľad je makroskopický – symbolický, ktorý operuje so symbolmi a s ich

transformáciou na iné symboly pomocou pravidiel. To znamená, že konekcionizmus nám poskytuje pojmový a argumentačný aparát na interpretáciu symbolického prístupu pomocou takých elementárnych pojmov, akými sú napr. aktivity jednotlivých neurónov, charakter spojov medzi neurónmi (amplifikačný alebo inhibičný) a pod. Podrobnosti o vzťahu medzi subsymbolickým a symbolickým prístupom v umelej inteligencii a v kognitívnej vede možno nájsť v literatúre [22].

1.2 Klasický konekcionizmus

Aj keď všeobecné rozšírenie konekcionizmu v umelej inteligencii a v kognitívnej vede je otázkou len ostatných 10-20-tich rokov, jeho korene možno sledovať vo filozofii už u Aristotela (citované podľa [30]). Ten v *koncepcii pamäti* predpokladá, že pamäť je zložená z jednoduchých elementov (pojmov, koncepcií, symbolov,...), ktoré sú navzájom spriahnuté (spojené) ich podobnosťou prostredníctvom rôznych mechanizmov (napr. časovou následnosťou, objektovou podobnosťou a susedstvom v priestore). Tieto asociované štruktúry môžu byť kombinované do zložitejších štruktúr tak, že vzniká usudzovanie a výber z pamäti. Táto Aristotelova idea veľmi pripomína súčasný symbolický prístup Newela a Simona [22] k algoritmizácii asociatívnej pamäti. Mnohé dôležité postuláty konekcionizmu sú už obsiahnuté v dielach materialistov (napr. la Mettrie a Hobbes) a taktiež v dielach britských empirikov (napr. Berkeley, Locke, Hume). Podľa materialistov platí, že neexistuje nič iné okrem hmoty a energie, všetky ľudské vlastnosti - vrátane myslenia – sa dajú vysvetliť výlučne pomocou fyzikálnych procesov prebiehajúcich v ľudskom tele, a najmä v mozgu. Toto viedlo empirikov k záveru, že ľudské poznanie možno odvodiť výlučne zo zmyslovej skúsenosti, a navyše, asociácie týchto skúseností vedú k mysleniu. To znamená, že ľudská kognícia sa riadi fyzikálnymi zákonmi a možno ju skúmať empiricky.

V druhej polovici 19. storočia mal konekcionistický prístup k interpretácii kognitívnych aktivít ľudského mozgu veľkú podporu v prírodovedne orientovanej psychológii a filozofii. Psychológovia Spencer [31] a James [13] vo svojich dielach uvádzajú príklady konekcionistických sietí, pomocou ktorých kombinovali základné princípy asociancizmu s dobovými neurologickými predstavami.



Obrázok 1.3. Konexie medzi aferentnými - vstupnými (A) a eferentnými - výstupnými (E) "konexiami", ktoré majú umožniť účinný prenos signálu pre pohyb svalu. Body *a* a *e* označujú miesta, kde nastáva divergencia "konexií".

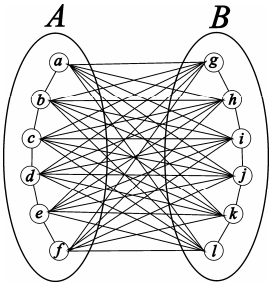
1.2.1 Spencerove "konexie"

Spencer sa zaslúžil o konečné oddelenie psychológie od filozofie. Jeden z jeho základných princípov použitých v jeho knihe z r. 1855 [31] bola idea, že pochopenie základných princípov fungovania nervového systému je nevyhnutným predpokladom pre korektnú interpretáciu procesov a fenoménov psychológie. Tejto problematike venoval niekoľko kapitol, v ktorých podrobne opisoval dobové predstavy o fungovaní nervových štruktúr. Pri rozvíjaní týchto predstáv vychádzal z predpokladu, že existuje principiálna väzba medzi nervovými zmenami a psychickými stavmi. Svoju pozornosť upriamil hlavne na opis vzniku a na význam spojov medzi neurónmi, pozri obr. 1.3.

Spencer bol prvým psychológom, ktorý si uvedomil úzky kauzálny vzťah medzi vonkajšími vzťahmi medzi objektmi prostredia, či ich atribútmi a vnútornými vzťahmi existujúcimi medzi nervovými bunkami: "ak pre nejaký interný stav *a* existuje tendencia, aby bol nasledovaný iným interným stavom *d*, potom väzba medzi nimi sa zosilňuje (alebo zoslabuje) v závislosti od toho, či *A* a *D* (vonkajšie objekty alebo atribúty, ktoré produkujú interné stavy *a* a *d*) sa vyskytujú spolu". Ak *D* je dôsledkom *A*, potom *interný* stav *a* indukuje *interný* stav *d*. Navyše, ak vonkajšie stavy *A* a *D* sa často vyskytujú spolu (v časovej alebo priestorovej súslednosti), potom vnútorné stavy *a* a *d* koexistujú. Ak medzi vonkajšími stavmi *A* a *D* existuje vzťah príčina - účinok, potom vnútorný stav *a* indukuje vnútorný stav *d*. Spencer uzaviera tieto úvahy formuláciou predpokladu, že sily spojov - konexií medzi internými stavmi, ktoré sú priradené externým udalostiam, sú dôležité a musia predstavovať hlavný objekt štúdia psychológie. Inými slovami, ***Spencer formuloval predpoklad, že vedomosti o externom svete sú zakódované v mozgu pomocou konexií.***

1.2.2 Jamesova asociatívna pamäť

James vo svojej dvojdielnej knihe "*The Principles of Psychology*" z r. 1980 [13], podobne ako Spencer, vychádzal z postulátu, že psychologické fenomény sa musia vysvetľovať pomocou aktivít mozgu, a existuje úplný paralelizmus medzi analýzou fungovania nervov a analýzou mentálnych ideí, každá mentálna modifikácia musí byť sprevádzaná telesnými zmenami. Najmarkantnejším príkladom konekcionizmu u Jamesa je jeho model asociatívnej pamäti, ktorý obsahuje jednotlivé idey, ktoré sú medzi sebou prepojené do paralelnej štruktúry tak, že vybavenie si (recall) jednej idey je sprevádzané súčasným pripomenutím si asociovaných ideí (pozri obr. 1.4).



Obrázok 1.4. Jamesov model distribuovanej pamäti. Aktivácia udalosti *A* spôsobuje aktiváciu udalosti *B* pomocou príslušných spojov ohodnotených váhami.

James si uvedomil skutočnosť, že ak sú si objekty veľmi podobné, potom aktivácia jedného z nich indukuje aktiváciu všetkých ostatných podobných objektov. Takáto neželateľná aktivácia celého obsahu pamäti (total recall), je obmedzená jeho pravidlom premenného záujmu: niektoré procesy prebiehajúce v mozgu majú vždy inú váhu ako iné procesy. Týmto vlastne James postuloval existenciu rôznych váh spojov v asociatívnej pamäti, čo vyvrcholilo vo formulácii jeho slávneho zákona (law of neural habit): "***ak dva elementárne procesy sa v mozgu vyskytujú súčasne (alebo sú bezprostredne následne aktivované), potom pri znovuopakovaní aktivácie jedného z týchto procesov vzniká tendencia, že tento proces bude excitovať druhý proces***" [31, vol. 1, str. 566]. Inými slovami, ak dve udalosti sa vyskytujú opakovane, potom váha ich spoja sa zvyšuje. Treba poznamenať, že James vo svojich úvahách o procesoch zosilnenia opakovane sa vyskytujúcich asociácií myslí fyzikálne procesy prebiehajúce v mozgu. Jeho vysvetlenie mechanizmu asociatívnej pamäti znamená pretransformovanie historicky prvého aristotelovského asociancizmu na biologický

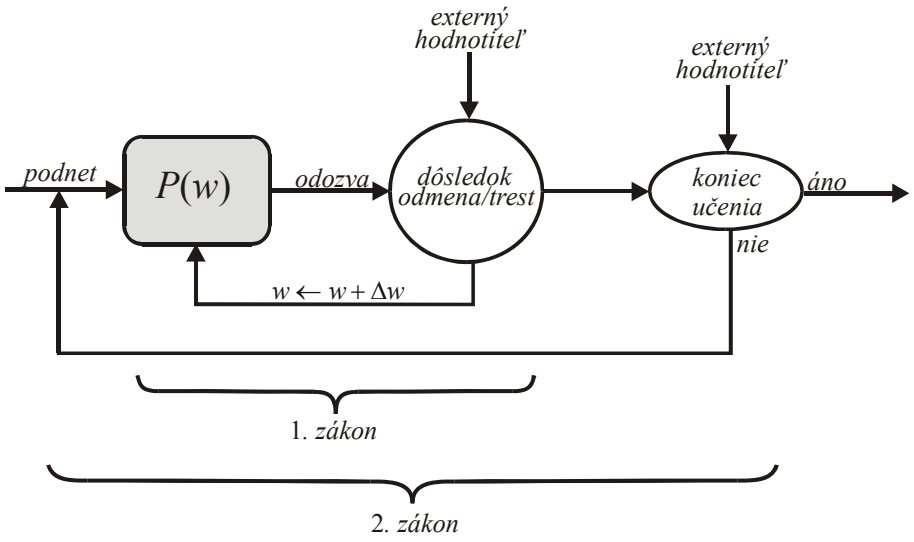
konekcionizmus, kde sa operuje s biologickými pojmami, akými sú neuróny, spoje medzi nimi a premenlivá intenzita týchto spojov.

1.2.3 Thorndikov konekcionizmus

Thorndike vo svojej knihe "*The Fundamentals of Learning*" z r. 1932 [32] ďalej tvorivým spôsobom rozvíjal myšlienky svojho učiteľa Jamesa. Aj keď jeho neskoršie práce už patria do obdobia 30-tych a 40-tych rokov dvadsiateho storočia, musíme jeho dielo pokladať za vyvrcholenie klasického konekcionizmu pestovaného a rozvíjaného v psychológii na prelome 19. a 20. storočia.

Thorndikov konekcionizmus [33] už dôsledne odlišuje sub-symbolický pohľad na neurálne asociácie od všeobecného "voľného" konekcionizmu svojho učiteľa Jamesa. Koncept neurálnych spojov sa stal fundamentálnym pre jeho konekcionizmus. Táto skutočnosť mu umožnila uskutočniť "algoritmický" opis učenia sa úplne pomocou tohto konceptu. Študoval adaptívne zmeny v správaní sa zvierat ako analógiu k ľudskému učeniu sa a rozpracoval myšlienku, že behaviorálne asociácie (connections - spoje) sú predikovateľné pomocou dvoch zákonov:

1. ***Zákon účinku* (*the law of effect*)** sa zaoberá pôsobením odmeny/trestu na opakujúce sa, bezprostredne po sebe idúce podnety (vstupy, stimuly) a odozvy (výstupy, reakcie). Ak je následok správania príjemný (odmena), tak sa neurálny spoj medzi podnetom a odozvou zosilňuje. V opačnom prípade, ak dôsledok na stimul je záporný (trest), potom väzba medzi podnetom a odozvou postupne zaniká.
2. Podľa ***zákona opakovaného používania* (*the law of exercise*)** je požadované správanie výsledkom častého používania dvojice podnet a odozva.



Obrázok 1.5. Formálna schéma Thorndikovho učenia sa "s odmenou a trestom". Mozog "žiaka" je reprezentovaný kognitívnym modulom $P(w)$, ktorý má určitú plasticitu (schopnosť meniť sa pomocou učenia sa) reprezentovanú "parametrom" w, od hodnoty ktorého závisí funkčnosť modulu. Podľa 1. zákona, integrálnou časťou učenia sa je hodnotenie externým učiteľom, ktorý podľa aktuálnej aktivity kognitívneho modulu rozhodne, či "žiak" bude odmenený alebo potrestaný, potom pomocou tohto rozhodnutia sa upraví kognitívny modul "žiaka" zmenou parametra $w \leftarrow w + \Delta w$. Pomocou 2. zákona sa vykoná zafixovanie požadovaného správania sa "žiaka" tým, že sa elementárny akt učenia mnoho ráz opakuje.

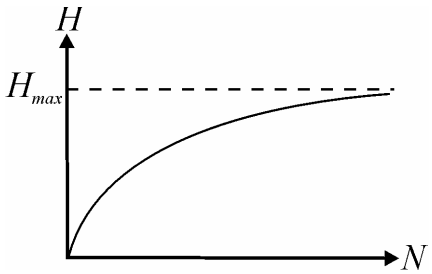
Učenie sa, ktoré je založené na týchto dvoch zákonoch, sa v informatike nazýva „***učenie sa s odmenou a trestom***“ („***reinforcement learning***“, pozri obr. 1.5). Formalizácia "učenia sa s odmenou a trestom" v prípade, že kognitívny modul "žiaka" je realizovaný doprednou neurónovou sieťou, tvorí teoretický základ širokej triedy konekcionistických metód pre učenie sa v zložitom a dynamicky meniacom sa prostredí [3].

1.2.4 Hullovo pravidlo učenia sa

Hull pomocou dobových predstáv o neurónoch navrhol v r. 1943 [12] niekoľko empiricky testovateľných rovníc o priebehu učenia sa. Východiskom jeho úvah bol predpoklad, že základom procesu učenia sa je zosilnenie váh vybraných neurónových spojov vzhľadom na ostatné neuróny, alebo vznik úplne nových spojov. Postuloval nasledujúce jednoduché vlastnosti neurónov:

(1) Výstupný neurónový impulz (s_1) je nelineárnou funkciou vstupného impulzu,
(2) interakcia medzi výstupnými neurónovými impulzmi (s_1 & s_2) znamená, že odozva na rovnaké podnety nemusí byť za každej situácie konštantná, a
(3) spontánny vznik nervového impulzu spôsobuje variabilitu správania za konštantných podmienok.

Podľa Hulla je učenie sa "výsostne dôležitý biologický proces", spočívajúci nielen v modifikovaní neurónových spojov vzhľadom na ostatné (konštantné) spoje, ale aj vo vzniku úplne nových spojov. Proces učenia sa Hull pokladá za úplne automatický, je výsledkom interakcie medzi organizmom a okolím. Ako ilustráciu Hullových úvah naznačíme jednoduché odvodenie jeho slávnej formuly vyjadrujúcej vzrast "***sily zvyku***" (***habit strength***) vzhľadom na počet krokov učenia sa. Nech premenná H vyjadruje silu zvyku, jej maximálna hodnota je označená H_{max}, a nech N je počet krokov učenia sa. Použijeme nasledujúcu fenomenologickú rovnicu, ktorá vyjadruje vzťah medzi silou zvyku a počtom krokov (pozri obr. 1.6)



Obrázok 1.6. Graf znázorňujúci závislosť sily zvyku H od počtu krokov N. Graf je monotónne rastúci a asymptoticky sa približuje k maximálnej hodnote H_{max} pre $N\to\infty$ (pozri rov. (7.2)).

$$H(N) = H_{max}\left(1 - e^{-\alpha N}\right) \tag{1.2}$$

kde α je kladná konštanta rýchlosti učenia sa. Po jednoduchých úpravách môžeme dostať výraz pre prírastok sily zvyku

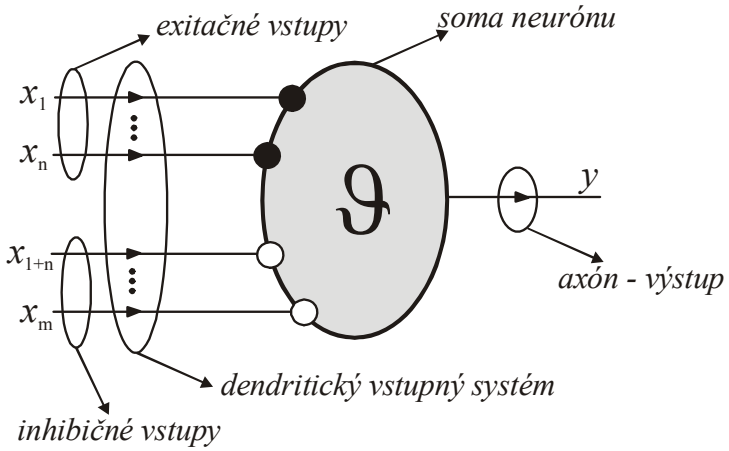
$$\Delta H = H(N+1) - H(N) = f \times \left(H_{max} - H(N)\right) \tag{1.3}$$

kde $f = 1 - e^{-\alpha}$ je konštanta určená pomocou rýchlosti učenia sa. Podľa formuly (1.3) pri opakovanom učení je prírastok sily zvyku úmerný rozdielu medzi maximálnou a aktuálnou silou zvyku. Táto formula nápadne pripomína formuly učenia sa známe z modernej teórie neurónových sietí.

1.2.5 McCullochova a Pittsova neurónová sieť

Ďalší významný príspevok do rozvoja konekcionizmu bol vykonaný neurofyziológom Warrenom McCullochom a vtedy 20-ročným študentom logiky Walterom Pittsom. Ich práca "*A logical calculus of the ideas immanent in nervous activity*", publikovaná v r. 1943 [19], znamená významný medzník v rozvoji konekcionizmu. V tejto práci už viac ako pred polstoročím sa s geniálnou jasnozrivosťou ukázalo, že neurónové siete sú mocným modelovým prostriedkom, napríklad, že ich siete sú schopné simulovať ľubovoľnú Boolovu funkciu.

Elementárnou jednotkou McCullochových a Pittsových neurónových sietí je ***logický neurón*** (výpočtová jednotka), pričom stav neurónu je binárny (t. j. má dva stavy, 1 alebo 0). Takýto logický neurón možno interpretovať ako jednoduché elektrické zariadenie - relé. Predpokladajme, že 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 y_1, y_2, ..., y_m, ktoré zoslabujú odozvu), pozri obr. 1.7.



Obrázok 1.7. Znázornenie McCullochovho a Pittsovho neurónu, ktorý obsahuje dendritický systém pre vstupné (excitačné alebo inhibičné) aktivity, axón pre výstup neurónovej aktivity. Soma (telo neurónu) je charakterizovaná prahovým koeficientom ϑ.

Aktivita logického neurónu je jednotková, ak suma excitačných vstupných aktivít mínus suma inhibičných aktivít je väčšia alebo rovná prahu $-\vartheta$, v opačnom prípade je nulová

$$z = \begin{cases} 1 & \left(x_1 + \ldots + x_n - y_1 - \ldots - y_m \geq -\vartheta\right) \\ 0 & \left(\text{v opačnom prípade}\right) \end{cases} \tag{1.4a}$$

Pomocou krokovej funkcie $s(\xi)$ definovanej (1.1b) môžeme aktivitu z vyjadriť takto

$$\begin{aligned} z &= s\left(x_1 + \ldots + x_n - y_1 - \ldots - y_m + \vartheta\right) \\ &= s\left(|x| - |y| + \vartheta\right) \end{aligned} \tag{1.4b}$$

kde $|x|=x_1+\ldots+x_n$ a $|y|=y_1+\ldots+y_n$. Tieto vzťahy pre aktivitu logického neurónu môžeme jednoducho interpretovať tak, že excitačné aktivity vstupujú do neurónu cez spoje, ktoré sú ohodnotené jednotkovým váhovým koeficientom (w=1), zatiaľ čo inhibičné aktivity vstupujú do neurónu cez spoje so záporným jednotkovým váhovým koeficientom (w=-1). Poznamenajme, že takto formulovaný logický neurón je bezprostrednou implementáciou všeobecných predstáv o fungovaní neurónu, ktoré boli formulované v úvodnej časti tejto kapitoly.

Logické neuróny sú schopné simulovať logické spojky, ktoré sú charakterizované ako lineárne separovateľné (napr. disjunkciu, konjunkciu, implikáciu a negáciu). Logické spojky, ktoré nie sú lineárne separovateľné (napr. ekvivalencia a exkluzívna disjunkcia XOR) nemôžu

byť simulované logickým neurónom. Táto skutočnosť naznačuje, že logický samotný neurón nie je univerzálne výpočtové zariadenie, existujú úlohy (napr. logická spojka XOR), ktoré nie sú riešiteľné pomocou logického neurónu.

Bolo dokázané už Mc Cullochom a Pittsom, že ľubovolná Boolova funkcia je simulovateľná pomocou neurónovej siete, ktorej vstupné neuróny špecifikujú premenné funkcie a výstupný logický neurón špecifikuje funkčnú hodnotu simulovanej Boolovej funkcie. Táto neurónová sieť už obsahuje tzv. skryté neuróny, ktoré spracuvávajú vstupné aktivity (pravdivostné hodnoty premenný Boolovej funkcie) tak, aby boli lineárne separovateľné pre výstupný neurón. Táto vlastnosť môže byť zosilnená tak, že neurónová sieť má univerzálny charakter trojvrstvovej neurónovej siete.

Veľkú zásluhu na pochopení a správnej interpretácii práce McCullocha a Pittsa má Minsky, ktorý vo svojej knihe "*Computation: Finite and Infinite Machines*" z r. 1967 [20] dôkladne analyzoval ich výsledky a vykonal niekoľko ďalších zovšeobecnení ich prístupu (najmä na konečno-stavové automaty). Navyše, štýl, akým bola písaná pôvodná práca McCullocha a Pittsa, nemá ďaleko od úplnej nezrozumiteľnosti. Minsky preformuloval ich výsledky do zrozumiteľnej formy a uviedol ich aj do rámca vtedajšej počítačovej vedy.

Je zaujímavé, že Pitts a McCulloch, teda tí istí autori, čo navrhli logický neurón, sa ďalej v roku 1947 začali zaoberať aj problémom, ako nervové systémy dokážu zvládnuť aj komplexnejšie úlohy z psychológie, ako je napr. videnie. Skúmali, ako sa rozpoznáva ten istý objekt, keď sa objaví v rôznych častiach zorného poľa, a ako je mozgový kmeň (superior colliculus) schopný transformovať priestorové zobrazenie senzorických vstupov do riadenia motorických aktivít, ako je pohyb oka. Opustili tak formálnu logiku v prospech výskumu priestorovej reprezentácie a analógových výpočtov, a týmto predznačili vývoj modelovania v neurovede na desaťročia vopred.

1.2.6 Hebbovo pravidlo učenia sa

Kniha kanadského neuropsychológa Donalda O. Hebba "*The Organization of Behaviour: A Neuropsychological Theory*" z r. [8] sa v súčasnosti pokladá za míľnik vo vývoji názorov na správanie sa jednotlivcov ako na problém pochopenia a interpretácie aktivít nervového systému (a naopak). Jeho obhajoba interdisciplinárneho prístupu k tomuto neuropsychologickému problému bola leitmotívom jeho celoživotného diela. Hebb bol prvý psychológ, ktorý systematicky používal termín "konekcionizmus" na označenie neurobiologického prístupu k interpretácii psychologických fenoménov. Hebbova kniha znamenala významný dobový "bod obratu" psychológie, ponúkla jednoduchý prístup k vysvetleniu fenoménov a javov psychológie pomocou "elementárnych" procesov a štruktúr obsiahnutých v mozgu:

(1) Spoje medzi neurónmi majú rastúcu účinnosť v závislosti od stupňa korelácie medzi pre- a postsynaptickou aktivitou. V neurovede je tento postulát známy ako Hebbova synapsia. Tento postulát môže byť preformulovaný do "pravidla učenia sa" v umelých neurónových sieťach.

(2) Skupiny neurónov, ktoré sú súčasne aktívne pri špecifickej udalosti, tvoria súbor neurónov, ktorý možno pokladať za reprezentanta danej udalosti.

(3) Myslenie je postupnou aktiváciou množiny súborov neurónov.

Formálne vyjadrenie pravidla učenia sa je pomerne jednoduché. Môžeme ho napríklad vyjadriť takto

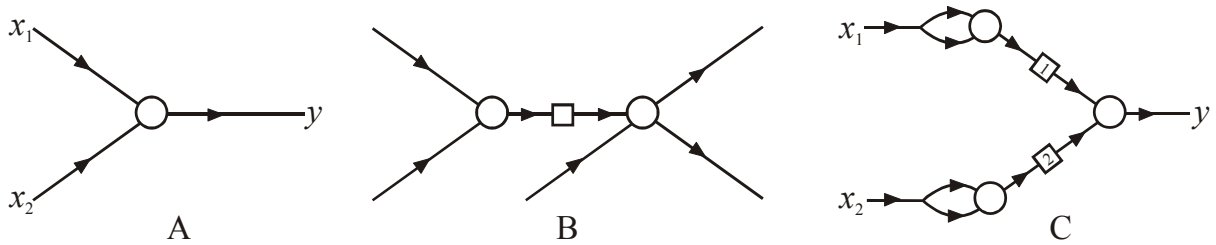
$$w_{ij}^{(t+1)} = w_{ij}^{(t)} + \alpha\, x_i^{(t)} x_j^{(t)} \tag{1.5}$$

kde $w_{ij}^{(t)}$ je ***váhový koeficient*** (synaptická sila) synapsie spájajúcej *i*-tý a *j*-tý neurón v čase *t*, $x_i^{(t)}$ a $x_j^{(t)}$ sú aktivity týchto neurónov v čase *t*, a $\alpha>0$ je parameter učenia sa (learning rate). Základnou formálnou nevýhodou tohto pravidla je, že môže viesť k váhovým koeficientom s nekonečnými hodnotami. Hebbovo pravidlo neprihliada na fakt, že hodnoty váhových koeficientov sú biologicky prijateľné len v určitom intervale prípustných hodnôt.

1. 2. 7 Turingov neorganizovaný stroj

Alan Turing v neopublikovanom rukopise "*Intelligent Machinery*" [34] z r. 1948 popisuje tzv. *neorganizovaný stroj*, ktorý možno chápať ako jednu z prvých špecifikácií umelej neurónovej sieti. Vtedajší riaditeľ *National Physical Laboratory* v Londýne Sir Charles Darwin (vnuk zakladateľa evolučnej teórie), kde bol Turing zamestnaný, nedoporučil rukopis publikovať v odbornom časopise, označil ho za "*schoolboy essay*" a "*smudgy*".

Turing definoval "***neorganizovaný stroj B-typu***" (dnes by sme povedali neurónovú sieť) ako výpočtovú štruktúru, ktorá obsahuje neuróny, spoje a modifikátory spojov. Každý neurón má dva vstupy a jeden alebo viac výstupov (ktoré môžu byť interpretované ako rozvetvený jeden výstup), pozri diagram A, obr. 1.8. Hodnoty výstupu *y* sú určené pomocou Boolovej funkcii NAND (v logike sa nazýva Shefferov symbol), pozri Tab. 1.1.



Obrázok 1.8. (A) Znázornenie neurónu, ktorý má dva vstupy x_1 a x_2 a jeden výstup *y*. (B) Každý spoj medzi dvoma neurónmi obsahuje modifikátor reprezentovaný štvorcom. (C) Jednoduchý Turingov neorganizovaný stroj – neurónová sieť, ktorý obsahuje dva vstupné neuróny a jeden výstupný neurón, pričom spoje medzi vstupným a výstupným neurónom sú riadené modifikátormi 1 a 2.

Každý spoj obsahuje modifikátor, ktorý môže byť v dvoch módoch:

(1) ***prechodový mód***, modifikátor neguje signál prechádzajúci spojom, t. j. 1 je zmenená na 0 a naopak,

(2) ***blokačný mód***, výstup z modifikátora je vždy signál 1 nezávisle na signálu vstupujúcom do modifikátora.

Tabuľka 1.1. Pravdivostné hodnoty funkcie NAND

x_1	x_2	y
0	0	1
0	1	1
1	0	1
1	1	0

Práca Turingovho neorganizovaného stroja (neurónovej siete) bude ilustrovaná jednoduchou sieťou znázornenou na obr. 1.9. Budeme predpokladať, že módu modifikátorov sú v štyroch

možných stavoch, výsledky sú sumarizované v Tab. 1.2, kde *mod* = 1(2) znamená, že príslušný modifátor je v prechodovom (blokačnom) móde.

Tabuľka 1.2. Výstupné aktivity neurónovej siete z obr. 1.8, diagram C, pre rôzne módy modifikátoirov

x_1	x_2	mod_1=1, mod_2=1	mod_1=1, mod_2=2	mod_1=2, mod_2=1	mod_1=2, mod_2=2
0	0	1	1	1	0
0	1	1	1	0	0
1	0	1	0	1	0
1	1	0	0	0	0

Prítomnosť modifikátorov na spojoch neutrónov umožňuje zaviesť proces učenia neurónovej siete, čo v Turingovej terminológii sa nazýva „***vhodná interferencia napodobňujúca vzdelávanie***“. Turing v práci taktiež teoretizoval o tom, že kortex dieťaťa je neorgizovaný stroj, ktorý môže byť „organizovaný“ pomocou vhodného tréningového učenia. Turing plánoval študovať zložitejšie modely kortextu pomocou jeho neorganizovaných strojov spôbom, ktorý je v súčasnosti bežný v modernom konekcionizme: simulovať umelé neurónové siete a ich učenie pomocou počítačov. Záverom rukopisu použil túto zaujimavú formuláciu: „*umožniť systému vývoj v priebehu určitého časového úseku, potom vykonať prerušenie vývoja ako inšpektor v škole a prekontrolovať vzniknutý progres*“. Žiaľ jeho výskum o neurónových sieťach nepokračoval, celé svoje pracovné úsilie venoval návrhu elektronického počítača schopného všeobecného použitia.

Záverom možno konštatovať, že napriek tomu, že Turingov neorganizovaný stroj z dnešného pohľadu reprezentuje exotickú neurónovú sieť, Turing bol zrejme jeden z prvých, ktorý hovoril explicitne o probléme učenia neurónovej sieti ako o probléme zmeny architektúry siete pomocou modifikátorov spojov. Žiaľ, ako robiť systematickým spôsobom tieto zmeny, aby sme dosiahli požadované vlastnosti na výstupe neorganizovaného stroja, Turing ani okrajovo nešpecifikoval, najskôr nemal o tom vôbec jasnú predstavu, ako to realizovať.

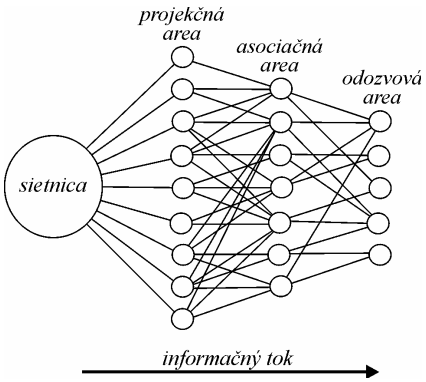
1.3 Počítačový konekcionizmus

1.3.1 Perceptrón

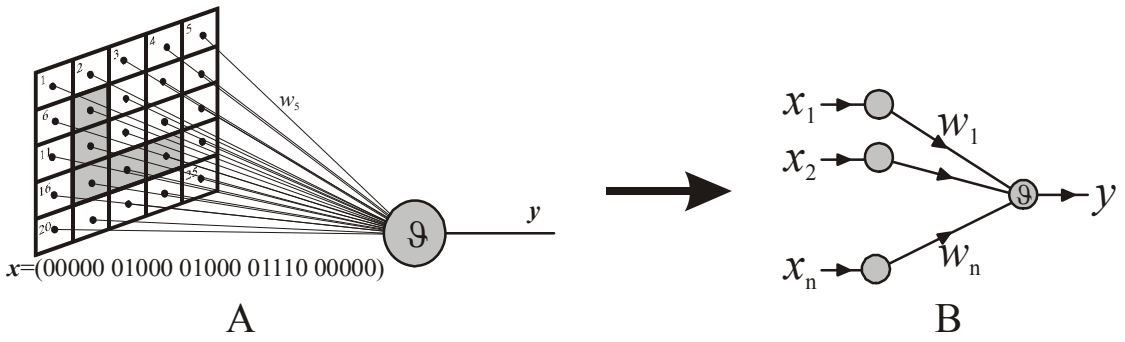
V predchádzajúcej časti tejto kapitoly sme sa zaoberali McCullochovými a Pittsovými neurónovými sieťami, ktoré obsahujú logické neuróny. Základnou črtou tohto prvého konekcionistického modelu je, že sily spojov (váhové koeficienty synapsií) sú nemenné, v týchto sieťach neexistuje učenie sa, ktoré by menilo váhové koeficienty. Je však známe, že aj napriek týmto dvom vážnym nedostatkom McCullochove a Pittsove neurónové siete sú efektívnym výpočtovým nástrojom, simulujú ľubovoľné Boolove funkcie a konečné automaty. Koncom 50-tych rokov psychológ Frank Rosenblatt opísal architektúru perceptrónu (pozri obr. 1.8), ktorú zanalyzoval vo svojej knihe "*Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*" [23]. Perceptrón bol zovšeobecnením logických McCullochových a Pittsových neurónov, podstatnou inováciou bol predpoklad existencie váhových koeficientov spojov (synapsií) a proces učenia sa, ktorým sa nastavujú tieto koeficienty tak, aby perceptrón korektne klasifikoval tréningovú množinu A_{train} zloženú z konečného počtu objektov $A_{train} = \{O_1, O_2, ..., O_n\}$.

Minsky a Papert (v knihe "*Perceptron: An Introduction to Computational Geometry*" z r. 1969) [21] pre potreby podrobnej analýzy výpočtových možností Rosenblattovho perceptrónu navrhli jeho abstrakciu, ktorá sa v súčasnej literatúre považuje za "pravý" perceptrón (pozri obr. 1.9). Výstupná aktivita je určená podobným spôsobom ako pre McCullochov a Pittsov logický neurón

.



Obrázok 1.8. Klasický perceptrón podľa Rosenblatta. Oblasť, kde sa premieta optickým systémom oka pozorovaný objekt, sa nazýva sietnica. Tá prenáša binárne hodnoty do vrstvy nazývanej projekčná oblasť (v lekárskej terminológii projekčná area), kde sa binárne kódovaný obraz numericky predspracováva. Spoje medzi sietnicou a projekčnou oblasťou sú pevné a neadaptabilné. Spoje do druhej vrstvy (asociačnej oblasti/arey) a tiež aj do tretej vrstvy (odozvová oblasť/area) sú stochasticky generované. Základným cieľom adaptačného procesu perceptrónu je nastaviť váhové koeficienty spojov tak, aby aktivity neurónov z tretej vrstvy (odozvová oblasť) správne klasifikovali obraz dopadajúci na sietnicu.



Obrázok 1.9. (A) Minského a Papertov perceptrón sa skladá z dvoch vrstiev. V prvej vrstve sú vstupné neuróny, pomocou ktorých sa kóduje obrázok na "sietnici" perceptrónu. V prípade, že daná štvorcová oblasť sietnice je biela, príslušná zložka binárneho vektora ***x*** je nulová; v opačnom prípade, keď daná oblasť je tmavá (je súčasťou premietnutého obrázka na sietnici), príslušná zložka v ***x*** je jednotková. Druhá vrstva obsahuje len jeden (výstupný) neurón, ktorého binárna aktivita kóduje písmeno zo sietnice oka. Ak je toto písmeno "L", potom požadovaná výstupná aktivita je jednotková, v opačnom prípade, pre všetky ostatné písmená je výstupná aktivita nulová. Každý spoj z *i*-tého vstupného neurónu do výstupného neurónu je ohodnotený váhovým koeficientom w_i. Výstupný neurón je ohodnotený prahom -ϑ. (B) Formálne vyjadrenie perceptrónu ako dvojvrstvovej neurónovej siete.

Podobne ako logický neurón, aj perceptrón je schopný *korektne klasifikovať* len takú množinu objektov A_{train}, ktorá je lineárne separovateľná. Z tejto dôležitej vlastnosti vyplýva, že výpočtové možnosti perceptrónu sú približne rovnaké ako McCullochovho a Pittsovho logického neurónu, aj keď sa v prípade perceptrónu požiadavka binárnosti vstupu môže vynechať a jednotlivé spoje sú ohodnotené váhovými koeficientami. Ani tieto dve zásadné

inovácie perceptrónu nestačia na odstránenie bariéry lineárnej separovateľnosti. Už Minsky s Papertom [21] ukázali, že táto bariéra môže byť prekonaná pomocou skrytých neurónov.

Perceptrón sa zásadne odlišuje od McCullochovho a Pittsovho logického neurónu tým, že pre perceptrón existuje jednoduchý algoritmus učenia sa. Algoritmus učenia sa perceptrónu bol formulovaný Rosenblattom pomocou modifikácie učenia sa navrhnutého Hebbom (pozri kapitolu 1.2.6). Rosenblatt tiež vypracoval dôkaz konvergencie tohto algoritmu (v priebehu konečného počtu krokov je schopný nájsť také nastavenie váhových koeficientov, ktoré vedie ku korektnej klasifikácii lineárne separovateľných objektov). Rosenblatt pred svojím predčasným skonom (mal iba málo vyše štyridsať) prišiel aj na existenciu obmedzení výpočtových schopností perceptrónu, navrhol rekurentné siete a viacvrstvový perceptrón, a skúmal možnosť samoorganizácie siete v rámci učenia sa bez učiteľa [23]. Dokonca vyrobil aj zariadenia s fotobunkami, asociátormi, so spojeniami implementovanými motoricky ovládanými potenciometrami, kde skúmal zašumené dáta a odpoveď pri porušení siete fyzickým odstránením drôtov, a pokúsil sa modelovať vizuálny systém mačky.

1.3.2 Kritika perceptrónu Minským a Papertom

Minsky koncom 60-tych rokov publikoval dve knihy: "*Computation: Finite and Infinite Machines*" [20] a "*Perceptron: An Introduction to Computational Geometry*" [21], ktorými sa zaslúžil o rozvoj teórie neurónových sietí. V knihe "*Computation*" vykonal autor dôkladnú analýzu práce McCullocha a Pittsa z roku 1943, ktorá sa jeho zásluhou stala známou v komunite informatikov zaoberajúcich sa neurónovými sieťami. Jedným z hlavných dôvodov tejto skutočnosti bolo, že štýl práce McCullocha a Pittsa je dobovo závislý a pre súčasného čitateľa takmer nezrozumiteľný. Zásluhou Minského bola táto práca preformulovaná do moderného jazyka a v niektorých aspektoch aj zovšeobecnená (napr. zaviedol pojem rekurentnej neurónovej siete, ktorá je schopná simulovať konečný automat).

Aj keď už na prelome 50-tych a 60-tych rokov bolo jasné, že perceptrón má určité obmedzenia, že nie je univerzálne použiteľný na klasifikáciu ľubovoľných obrazcov, až pomocou knihy "*Perceptron*" boli tieto obmedzenia presne opísané a špecifikované. Ukázalo sa a na rôznych príkladoch všeobecného charakteru sa demonštrovalo, čo dokáže perceptrón korektne klasifikovať a čo nie. Minsky a Papert zaviedli fundamentálny pojem *lineárnej separovateľnosti*, pomocou ktorého sa pomerne jednoducho (a tiež názorne) formulujú podmienky pre korektnú klasifikáciu danej triedy objektov. V dobe publikovania knihy bol jedným z najaktuálnejších problémov umelej inteligencie problém rozpoznávania a klasifikácie scén. Práve v tejto dobe vznikali základné "symbolické" algoritmy a postupy, ako riešiť problém analýzy a klasifikácie scén. Preto Minsky a Papert upriamili svoju pozornosť na riešenie týchto problémov pomocou perceptrónov. Zistili napríklad, že perceptrón je schopný rozlíšiť trojuholníky od štvorcov, ale nedokáže rozlíšiť súvislé geometrické telesá od nesúvislých. Neprekonateľné problémy má perceptrón aj s korektnou klasifikáciou nových obrazcov, ktoré vnikli z pôvodného obrazca aplikovaním translácie alebo rotácie. Ich záver sa dal preto očakávať: *perceptrón nie je univerzálne výpočtové zariadenie, existujú jednoduché úlohy, ktoré nie sú korektne klasifikovateľné perceptrónom*. V knihe boli navrhnuté aj rôzne zovšeobecnenia perceptrónu, a to použitie skrytých neurónov a neurónov vyššieho rádu. Autori ukázali, že tieto zovšeobecnenia perceptrónu sú schopné prekonať bariéru lineárnej separovateľnosti. Žiaľ, oba tieto prístupy boli spochybnené samými autormi tým, že nepovažovali za potrebné študovať metódu učenia sa pre nastavenie váhových koeficientov skrytých neurónov alebo neurónov vyšších rádov. Hlavné príspevky Minského a Paperta k teórii neurónových sietí možno zhrnúť takto:

(1) Minsky v knihe "*Computation: Finite and Infinite Machines*" pri štúdiu McCullochových a Pittsových neurónov zaviedol pojem rekurentnej neurónovej siete, ktorá bola zložená z prahových neurónov. Pomocou konštruktívneho dôkazu ukázal, že táto sieť simuluje konečný binárny automat. Podobnú problematiku, lenže na kvalitatívne vyššej úrovni, začali študovať až na prelome 80-tych a 90-tych rokov pomocou dynamických prístupov k teórii rekurentných neurónových sietí. Tieto siete sa chápu ako dynamické systémy so zložitými trajektóriami v stavovom priestore, ktoré môžu simulovať konečné automaty.

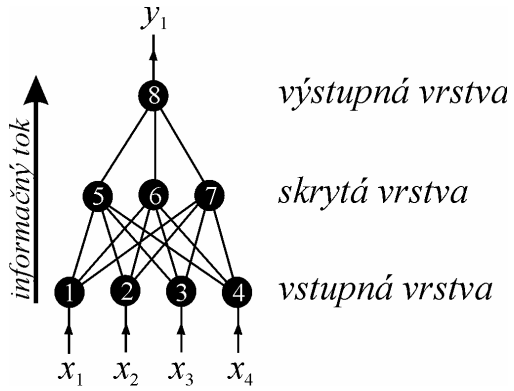
(2) Minsky a Papert v knihe "*Perceptron: An Introduction to Computational Geometry*" podrobne študovali výpočtové možnosti perceptrónu. Ukázali, že je schopný klasifikovať korektne len takú množinu objektov, ktorá je v zodpovedajúcom priestore vstupných aktivít lineárne separovateľná. Navrhli tiež dva možné prístupy, ako prekonať túto bariéru, a to buď použitím skrytých neurónov, alebo použitím neurónov vyšších rádov. Pretože nevenovali dostatočnú pozornosť procesu učenia sa takto zovšeobecnených perceptrónov, uvedené prístupy označili za neefektívne, pretože nie je k dispozícii metóda učenia sa, ktorá by určila hodnoty váhových koeficientov priradených týmto neurónom.

Prečo sa venujeme tak podrobne vyčísleniu zásluh týchto dvoch informatikov pre rozvoj neurónových sietí? V literatúre o neurónových sieťach sa už notoricky opakuje tvrdenie, že Minsky s Papertom svojou druhou knihou "*Perceptron*" spôsobili, že nastal na niekoľko desaťročí útlm výskumu v oblasti neurónových sietí. Podľa mienky autorov tejto kapitoly je tento názor nepravdivý, ba až smiešny. Táto vynikajúca kniha je stále primárnym zdrojom inšpirácie pre teóriu neurónových sietí a isto nespôsobila dlhodobý útlm rozvoja neurónových sietí. Hlavným zdrojom tohto útlmu bola skutočnosť, že vnútorné zdroje symbolickej umelej inteligencie neboli ešte zďaleka vyčerpané. Symbolická umelá inteligencia sa v tej dobe javila ako hlavný generátor nových trendov umelej inteligencie. Ďalším dôvodom bol odklon od problematiky rozpoznávania vzorov, v ktorom boli neurónové siete vtedy najlepšie, k problematike vyšších kognitívnych procesov, ako je napríklad riešenie problémov a dedukcia. Vtedy sa zdalo, že prostriedky klasickej umelej inteligencie dokážu tieto úlohy z vyšších kognitívnych procesov dobre riešiť. K tomu sa pridalo rozčarovanie z počiatočných neopodstatnených nádejí, keď priaznivci neurónových sietí, podobne ako ostatní vedci pracujúci v umelej inteligencii, prisudzovali vtedajším neurónovým sieťam aj kvality, ktoré tieto jednoducho nemali - čo sa skoro ukázalo. No aj tak boli neurónové siete aj naďalej rozvíjané hŕstkou nadšencov, ako bol napríklad Teuvo Kohonen [14,15] a Stephen Grossberg [6]. Situácia sa diametrálne zmenila až na prelome 80-tych a 90-tych rokov, keď si informatici začali čoraz naliehavejšie uvedomovať obmedzenosť metód symbolickej umelej inteligencie. Nepodarilo sa úplne uskutočniť žiadnu veľkú víziu symbolickej umelej inteligencie zo 60-tych a 70-tych rokov (napr. prekladač z prirodzeného jazyka, komunikácia s počítačom v prirodzenom jazyku, program pre šach na veľmajstrovskej úrovni,...).

1.3.3 Viacvrstvová neurónová sieť s dopredným šírením signálu

Na začiatku osemdesiatych rokov nastalo oživenie záujmu o neurónové siete. Čiastočne to bolo spôsobené neschopnosťou symbolických modelov klasickej umelej inteligencie postupne strácať funkčnosť s čiastočným poškodením podsystémov, obmedzenou schopnosťou generalizovať pre nové prípady a principiálnou neschopnosťou modelov vzrastať v zložitosti. V r. 1981 bol publikovaný zborník *Parallel Models of Associative Memory* [9], editovaný Geoffreyom Hintonom a Jamesom Andersenom. Ďalej John Hopfield, vynikajúci fyzik, v roku 1982 [10] opísal možnosti výpočtov neurónovými sieťami. Za začiatok moderného konekcionizmu možno považovať vydanie slávneho dvojdielneho zborníka "*Parallel Distributed Processing: Explorations in the Microstructure of Cognition*" [25]. V tomto zborníku boli publikované práce, ktoré už veľmi konkrétnym spôsobom (spolu s ich implementáciou na počítači) diskutovali vlastnosti neurónových sietí rôznych typov,

teoretické závery boli ilustrované teoretickými analýzami a počítačovými simuláciami. Medzi najdôležitejšie kapitoly v tomto zborníku isto patrí práca Rumelharta, Hintona a Williamsa [26] "*Learning Internal Representation by Error Propagation*" (ktorá bola publikovaná aj nezávisle v časopise Nature [27]), v ktorej bola opísaná metóda učenia sa neurónových sietí obsahujúcich skryté neuróny, pričom aktivačné funkcie neurónov boli diferencovateľné sigmoidy. Táto metóda je založená na minimalizácii účelovej funkcie typu, ktorá je spojitá a diferencovateľná. Použitím štandardnej gradientovej optimalizačnej metódy najprudšieho spádu (steepest descent) sa učenie sa neurónovej siete realizuje prostredníctvom minimalizácie účelovej funkcie. Architektúra 3-vrstvovej neurónovej siete je znázornená na obr. 1.10. Na rozdiel od perceptrónu, táto sieť už obsahuje skryté neuróny, ktoré ležia medzi vstupnými a výstupnými neurónmi.



Obrázok 1.10. Typická architektúra 3-vrstvovej neurónovej siete. Sieť obsahuje tri vrstvy: vstupnú vrstvu, ktorá je zložená zo vstupných neurónov (tieto nevykonávajú žiadne výpočty, ich význam je formálny, pomocou nich sieť prijíma vonkajšiu informáciu, vstupné aktivity x_1, x_2, x_3, x_4). Ďalšia vrstva obsahuje skryté neuróny, ktoré sú spojené všetkými možnými spôsobmi so vstupnými neurónmi. Horná vrstva obsahuje výstupný neurón, ktorého aktivita y_1 reprezentuje výstup siete. Výstupný neurón je spojený so skrytými neurónmi všetkými možnými spôsobmi. Každý spoj *i-j* je ohodnotený váhovým koeficientom w_{ij}, podobne, každý skrytý alebo výstupný neurón *i* je ohodnotený prahom ϑ_i.

Efektívnosť neurónovej siete pre klasifikáciu objektov z tréningovej množiny $A_{train} = \{O_1, O_2, ..., O_k\}$ je určená pomocou účelovej funkcie

$$E(O; w, \vartheta) = \frac{1}{2}\left(y_{req}(O) - y(O)\right)^2 \tag{1.6}$$

Ako optimalizovať túto účelovú funkciu pre neurónovú sieť, ktorá obsahuje skryté neuróny? Táto otázka má v histórii neurónových sietí principiálny charakter. Minsky s Papertom, aj keď si explicitne uvedomili dôležitosť riešenia tohto problému, sa ani nepokúsili o jeho riešenie. Azda hlavný formálny problém, na ktorý narazili Minsky s Papertom, bola nespojitá kroková aktivačná funkcia $s(\xi)$. Potom aj účelová funkcia (1.1b) je nespojitá, čiže štandardné gradientové metódy sú nepoužiteľné. Rumelhart so spolupracovníkmi obišli tento problém tak, že krokovú aktivačnú funkciu nahradili spojitou a diferencovateľnou sigmoidou

$$\sigma(\xi) = \frac{1}{1+e^{-\xi}} \tag{1.7}$$

Potom účelová funkcia (1.6) je diferencovateľná, čiže gradientové optimalizačné metódy sú úplne aplikovateľné. K výpočtu parciálnych derivácií účelovej funkcie (1.6) použil Rumelhart a spol. [26,27] jednoduchú metódu postupného použitia matematického pravidla o výpočte parciálnej derivácie zloženej funkcie.

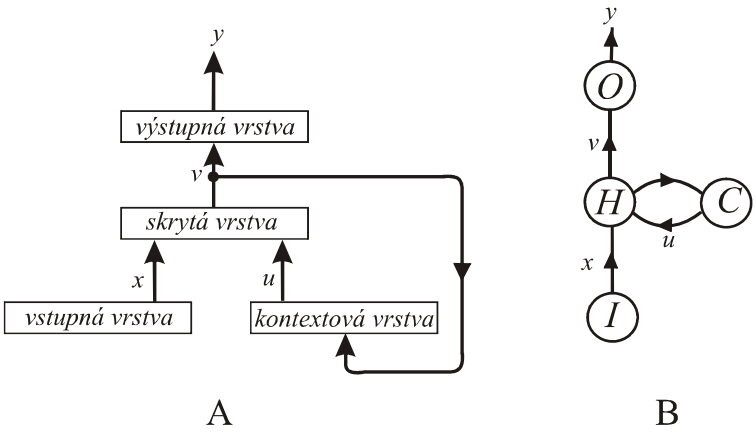
Vo všeobecnosti môžeme konštatovať, že neurónové siete sú schopné naučiť sa ľubovoľnú nekonfliktnú úlohu (napr. takú, ktorá vo svojej tréningovej množine nemá také dva rovnaké objekty, ktoré sú klasifikované rôzne). Môžeme si napr. vymyslieť ľubovoľnú

Boolovu funkciu. Táto "príjemná" skutočnosť našla svoj odraz aj v matematickej teórii neurónových sietí [11], kde sa dokázalo, že dopredné neurónové siete s jednou vrstvou skrytých neurónov sú univerzálnym aproximátorom; inak povedané, neurónové siete sú schopné naučiť sa ľubovoľnú úlohu reprezentovanú podobnou tabuľkou, akou bola vyjadrená úloha o sčítaní dvoch binárnych čísel. Žiaľ, táto matematická vlastnosť neurónových sietí je len existenčného charakteru, nešpecifikuje nám, koľko musí mať skrytých neurónov a aké sú hodnoty váhových koeficientov. Prijateľné hodnoty týchto parametrov musí zadať riešiteľ.

V začiatkoch rozvoja umelých neurónových sietí (koniec 80-tych rokov), keď nebolo známe, že neurónové siete sú univerzálnym aproximátorom, mnohé publikácie o neurónových sieťach boli zamerané na demonštráciu faktu, že neurónové siete sú schopné korektne klasifikovať ďalšie a zložitejšie nové úlohy. Po dôkaze vlastnosti „univerzálneho aproximátora“ na prelome 80-90-tych rokov, výskyt publikácií takéhoto charakteru náhle prestal. Poznatok, že dopredné siete sú schopné naučiť sa ľubovoľnú úlohu, stal sa bežným poznatkom teórie umelých neurónových sietí.

1.3.4 Rekurentné neurónové siete

Rekurentné neurónové siete boli prvýkrát diskutované už Minským [20] koncom 60-tych rokov pri jeho zovšeobecnení McCullochovho a Pittsovho logického neurónu. Ukázal, že rekurentná sieť (obsahujúca orientované cykly) je schopná simulovať konečný binárny automat. V tejto časti našej kapitoly budeme diskutovať o všeobecných vlastnostiach rekurentných sietí, zložených z neurónov so sigmoidnou aktivačnou funkciou. *Rekurentné siete sa budú chápať ako také neurónové siete, ktoré obsahujú orientované cykly* zo spojov medzi neurónmi, pozri obr. 1.11.



Obrázok 1.11. Elmanova rekurentná sieť. (A) Táto sieť vzniká jednoduchou modifikáciou štandardnej 3-vrstvovej doprednej siete zavedením tzv. kontextovej vrstvy, ktorá je časťou orientovaného cyklu. (B) Schematická reprezentácia Elmanovej rekurentnej siete [4,5], medzi skrytou a kontextovou vrstvou existuje orientovaný cyklus.

Aký je rozdiel medzi doprednou sieťou a rekurentnou sieťou? Dopredná sieť umožňuje jednoduchý postupný "dopredný" výpočet aktivít neurónov, pozri rovnice (7.17-18). Žiaľ, jednoduchý "dopredný" postup pre výpočet aktivít neurónov rekurentných sietí nie je možný. V dôsledku existencie orientovaných cyklov pri výpočte aktivít neurónov nastáva situácia, keď potrebujeme poznať aktivitu neurónu, ktorá ešte nebola v predchádzajúcom procese vypočítaná. Vo všeobecnosti môžeme povedať, že aktivity skrytých a výstupných neurónov z rekurentných sietí sú určené spriahnutým systémom nelineárnych rovníc. Takýto systém

rovníc môžeme riešiť iteračnou metódou, ktorá je formálne vyjadrená pomocou nasledujúcich rekurentných obnovovacích pravidiel

$$x_i^{(t+1)} = f_i\left(x_1^{(t)}, x_2^{(t)}, ..., x_n^{(t)}\right) \quad (i = 1, 2, ..., n) \tag{1.8}$$

kde f_i je nelineárna funkcia určujúca i-tú aktivitu. Proces opakovaného dosadzovania sa začína dosadením počiatočných hodnôt $x_1^{(0)}, x_2^{(0)}, ..., x_n^{(0)}$.

Základná vlastnosť rekurentných sietí je schopnosť rozoznávať sekvenciu znakov v postupnosti symbolov, čo má principiálny význam v informatike význam pre spracovanie časových radov a v kognitívnej vede pre modelovanie kognitívnych procesov súvisiacich s jazykom. Rekurentné siete sú vhodným konekcionistickým prostriedkom na klasifikáciu reťazcov premenlivej dĺžky. Tieto neurónové siete sú adaptovateľné tak, že sú schopné indikovať prítomnosť podreťazcov alebo iných charakteristík, ktoré majú význam v procese sémantickej interpretácie týchto reťazcov. Navyše, rekurentné neurónové siete sú schopné klasifikovať nielen lineárne reťazce znakov, ale aj zložitejšie štruktúry, akými sú napr. stromové syntaktické štruktúry, alebo vo všeobecnosti, acyklické grafy s ohodnotenými vrcholmi a/alebo hranami [4,5].

1.4 Zhrnutie

Hlavným cieľom tejto kapitoly bolo formulovať a ilustrovať základné princípy neurónových sietí (konekcionizmu), poukázať na jeho historické korene a zdroje, a tiež na hranice tohto prístupu. Problém pozície konekcionizmu v rámci výpočtovo orientovanej umelej inteligencie je neobyčajne zložitý. Do popredia vystupuje problém univerzálnej aplikovateľnosti neurónových sietí v umelj inteligencii, či je tento prístup vhodný na štúdium všetkých typov kognitívnych aktivít študovaných umelou inteligenciou a kognitívnou vedou. Ako preklenúť veľmi redukcionistický charakter modelovania pomocou neurónových sietí, použitých na štúdium vyšších kognitívnych aktivít? Súčasná umelá inteligencia ponúka kompromisné riešenie nazývané *hybridný prístup* [35]. Tento moderný prístup sa usiluje preklenúť často spomínaný "antagonizmus" medzi symbolizmom a konekcionizmom v umelej inteligencii tak, že hľadá spojenie symbolického a konekcionistického prístupu. Inak povedané, to, čo je dobre implementovateľné pomocou symbolického prístupu (napr. dlhodobé plánovanie), bude sa realizovať klasickými symbolickými metódami, a naopak, to, čo je dobre implementovateľné pomocou neurónových sietí (napr. rekognoskácia prostredia alebo reflexívne uvažovanie), bude sa realizovať pomocou neurónových sietí.

Musíme konštatovať, že sme zatiaľ ešte ďaleko od kompletného porozumenia spracovania informácie v ľudskom mozgu [2] pomocou neurónových sietí. Dokonca nie je celkom jasné ani to, čo je pre spracovanie signálov neurónov to najpodstatnejšie - či synchronizácia, hierarchické subštruktúry, alebo ešte niečo iné. Ďalej, niektoré konekcionistické modely obsahujú procedúry, najmä algoritmy učenia sa, ktoré sa s najväčšou pravdepodobnosťou v mozgu neodohrávajú. Biologická vierohodnosť modelov pri mnohých prístupoch ašpirujúcich na vysvetlenie vyšších kognitívnych aktivít, žiaľ, chýba. Aj keď pri procesoch úzko spojených s vnímaním máme pomerne presné modely založené na neurologických poznatkoch, o súčasných konekcionistických modeloch na pokročilejšie spracovanie informácie ("rozmýšľanie") hovoríme, že sú inšpirované biologickými neurónovými sieťami, a nie že sú ich vernou simuláciou. Musíme však konštatovať, že v inžinierky orientovaných aplikáciách výpočtovej inteligencie, neurónové siete poskytujú univerzálne aplikovateľný formalizmus pre zabezpečenie jednoduchej možnosti učenia sa daného inteligentného systému.

Cvičenia

Cvičenie 1.1. Aktivita neurónu v rámci konekcionistického modelu špecifikovaného v kapitole 1.1 je určená formulami (1.1a-b) (pozri taktiež obr. 1.2). Definujme štyri rôzne aktivity pomocou formúl

$$y_1 = s\left(x_1 + x_2 + 1\right)$$
$$y_2 = s\left(-x_1 + x_2 + 1\right)$$
$$y_3 = s\left(x_1 - x_2 + 1\right)$$
$$y_4 = s\left(x_1 + x_2 - 1\right)$$

Diskutujte tieto formule, pre ktoré hodnoty x_1 a x_2 poskytujú jednotkovú aktivitu a pre ktoré nulovú aktivitu, znázornite tieto oblasti v stavovom priestore x_1-x_2. Nakreslite priebehy týchto funkcií v 3D grafike pomocou nejakého počítačového systému (MATLAB, Mathematica alebo Maple).

Cvičenie 1.2. V kapitole 1.2.3 sú formulované základné princípy učenia s odmenou a trestom (pozri obr. 1.5). Predpokladajme, že pomocou tohto učenia agent je trénovaný na zvládnutie úlohy *U* (napr. pozdraviť dekana fakulty). Stav naučenia tejto úlohy nech je špecifikovaný parametrom $0 = w_{min} < w < 1 = w_{max}$, pričom počiatočná hodnota tohto parametru nech je $w^{(0)} = 0$. Pravdepodobnosť učenia je generovaná kvázináhodne tak, že je úmerná veličine *w*

$$p_{odmena} = \begin{cases} 1 & \left(random(0,1) < w\right) \\ 0 & \left(ináč\right) \end{cases}$$

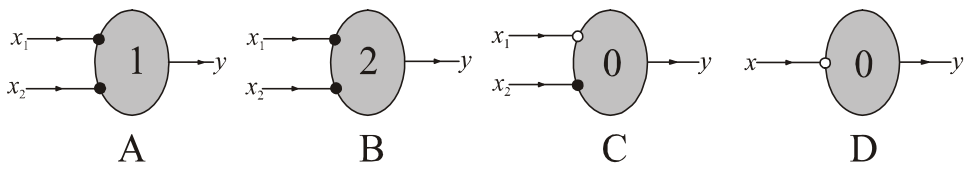
Kde *random*(0,1) je náhodné číslo s rovnomernou distribúciou a z otvoreného intervalu. Prevdepodobnosť trestu P_{trest} získame ako doplnok k pravdepodobnosti odmeny, $P_{trest} = 1 - P_{odmena}$. Ak pri výpočte P_{odmena} dostaneme, že je nulová, potom automaticky P_{trest} = 1 (a naopak). Oprava Δw nech je určená formulou

$$\Delta w = \begin{cases} \lambda\left(w_{max} - w\right) & \left(pre\ P_{odmena} = 1\right) \\ -\lambda w\left(w_{max} - w\right) & \left(pre\ P_{trest} = 1\right) \end{cases}$$

Kde $\lambda > 0$ je kladný parameter „rýchlosť učenia“. Úloha tohto príkladu spočíva v napísaní jednoduchého programu pre tieto formuly a simulácii priebehu učenia s odmenou a trestom a v grafickej vizualizácii výsledkov pre rôzne hodnoty parametra λ.

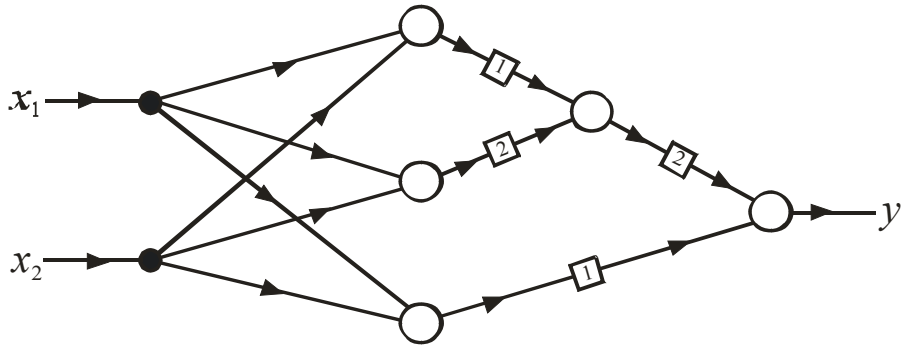
Cvičenie 1.3.. Navrhnite modifikáciu formuly (1.2) Hulovho učenia sa tak, aby obsahovalo aj proces zabúdania.

Cvičenie 1.4. Zostrojte tabuľky Boolových funkcií, ktoré sú reprezentované logickými neurónmi na obr. 1.12.



Obrázok 1.12. Rôzne tvary logického neurónu pre príklad 1.4, je potrebné zostrojiť tabuľky Boolových funkcií, ktoré sú špecifikované týmito neurónmi.

Cvičenie 1.5. Na obrázku 1.13 je znázornený Turingov neorganizovaný stroj, ktorý obsahuje dva vstupy binárnych aktivít x_1 a x_2, jeden binárny výstup y a štyri modifikátory. Preskúmajte všetkých $2^4 = 16$ neorganizovaných strojov pre všetky možné hodnoty modifikátorov a zostrojte tabuľku zobrazení binárnych vstupov (x_1, x_2) na výstup y.



Obrázok 1.13. Turingov neorganizovaný stroj, ktorý obsahuje štyri neuróny a štyri modihikátory. Pre dané hodnoty modifikátorov, je možné interpretovať toto výpočtové zariadenie ako zobrazenie $F : \{0,1\}^2 \xrightarrow{\{1,2\}^4} \{0,1\}$, kde nad šipkou je uvedená štvorica modifikátorov z karteziánskeho produktu $\{1,2\}^4$.

Cvičenie 1.6. Vyšetrite pomocou prostriedkov matematickej analýzy (t. j. pomocou vlastností prvej a druhej derivácie) priebeh sigmoidy, ktorej analytický tvar je špecifikovaný (1.7), nakreslite kvalitatívny graf tejto funkcie.

Cvičenie 1.7. Nech sigmoida (1.7) je modifikovaná pomocou kladného parametru $\alpha > 0$ takto

$$\sigma_\alpha(\xi) = \frac{1}{1+e^{-\alpha\xi}}$$

Zistite, aký má priebeh táto funkcia pre limitnú hodnotu parametra $\alpha \to \infty$ pre kladné, nulové a záporné hodnoty argumentu ξ

$$\lim_{\alpha\to\infty} \sigma_\alpha(\xi) = \begin{cases} ? & (\text{pre } \xi>0) \\ ? & (\text{pre } \xi=0) \\ ? & (\text{pre } \xi<0) \end{cases}$$

Literatúra

[1] Anderson, J. A., Pellionisz, A., and Rosenfeld, E (eds.): *Neurocomputing* 2. MIT Press, Cambridge, MA, 1990.

[2] Arbib, M.A. (ed.): *The Handbook of Brain Theory and Neural Networks.* MIT Press, Cambridge, MA, 1998.

[3] Barto, A. G. and Sutton, R. S.: *Reinforcement Learning: An Introduction. MIT Pr*ess, Cambridge, MA, 1998.

[4] Elman, J. L.: Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning* **7** (1991) 195-225.

[5] Elman, J.L., Bates, J.L., Johnson, M.H., Karmiloff-Smith, A, Parisi, D, and Plunkett, K.: *Rethinking Innateness : A Connectionist Perspective on Development***.** MIT Press, Bradford Books, Cambridge, MA, 1998.

[6] Grossberg, S.: Adaptive pattern classification and universal recoding: II. Feedback, expectation, olfaction, illusions. *Biological Cybernetics* **23** (1976) 187-202.

[7] Havel I.M.: Přirozené a umělé myšlení jako filozofický problém. V: Mařík V., Štěpánková O. a Lažanský J. (editori): *Umělá inteligence (3)*. Academia, Praha, 2001.

[8] Hebb, D. O.: *The organisation of behavior: a neurophysiological theory*. Wiley, New York, 1949.

[9] Hinton, G. E. and Anderson, J. A. (eds): *Parallel Models of Associative Memory*, Lawrence Erlbaum Associates, Publishers, Hillsdale, NJ, 1981.

[10] Hopfield, J. J.: *Neural networks and physical systems with emergent collective computational abilities*. *Proc. Natl. Acad. Sci. USA* **79** (1982) 2554-2558.

[11] Hornik, K., Stinchcombe, M., and White, H.: Multilayer feedforward networks are universal approximators. *Neural Networks* **2** (1989) 359-366.

[12] Hull, A.: *Principles of Behaviour*. Applenton-Century-Croft, New York, 1943.

[13] James, W.: *The Principles of Psychology*, vol. 1 and 2. Dover, New York, 1890.

[14] Kohonen, T.: Self-organized formation of topologically correct feature maps. *Biological Cybernetics* **43** (1982) 59-69.

[15] Kohonen, T.: *Self-Organizing Maps*. Springer-Verlag, Berlin, 1995.

[16] Kvasnička, V., Beňušková, Ľ., Farkaš, I., Kráľ, A., Pospíchal, J. a Tiňo, P.: *Úvod do teórie neurónových sietí*. IRIS, Bratislava, 1997.

[17] Kvasnička, V., Pospíchal, J.: Kognitívne vedy a konekcionizmus. In *Kognitívne vedy*, Rybár, J., Beňušková Ľ., Kvasnička, V. (ed.), Kalligram, Bratislava, 2002.

[18] McClelland, J. L. and Rumelhart, D. E.: An interactive activation model of context effects in letter perception. *Psychological Review* **88** (1981) 375-407.

[19] McCulloch,W. S. and Pitts, W.: A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* **5** (1943) 115-133.

[20] Minsky, M. L.: *Computation. Finite and Infinite Machines*. Prentice-Hall, Englewood Cliffs, NJ, 1967.

[21] Minsky, M. and Papert, S.: *Perceptrons. An Introduction to Computational Geometry*. MIT Press, Cambridge, MA, 1969.

[22] Newell, A. and Simon, D. A.: Computer science as emprical enquiry: symbols and search. *Communications of the ACM* **12** (1976) 113-126.

[23] Rosenblatt, F.: *Principles of Neurodynamics: Perceptrons and the Theory of Brain Machines*. Spartan Books, Washington, 1962.

[24] Rosenblatt, F.: The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review* **65** (1958) 386-408.

[25] Rumelhart, D. E. and McClelland, J. L.: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1-2. MIT Press, Cambridge, MA, 1986.

[26] Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Learning Internal Representations by Error Propagation. In: Rumelhart, D. E. and McClelland, J. L. (eds.): *Parallel Distributed Processing: Explorations in the Microstructure of Cognition (Vol. 1)*. MIT Press, Cambridge, MA, 1986.

[27] Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Learning representations by back-propagating errors. *Nature*, **323**(1986), 533-536.

[28] Siegelmann, H. T. and Sontag, E. D.: Analog computation via neural networks. *Theoretical Computer Science* **131** (1991) 331-360.

[29] Sontag, E. D.: Automata and neural networks. *The Handbook of Brain Theory and Neural Networks*, M. A. Arbib, eds. MIT Press, Cambridge, MA, 1995.

[30] Sorabji, R.: *Aristotelés o paměti* (preložil M. Pokorný). Rezek, Praha, 1995.

[31] Spencer, H.: *The Principles of Psychology*, Vol. 1 and 2. D. Apleton and Company, New York, 1855.

[32] Thorndike, E. L.: *The Fundamentals of Learning*. Teachers College, Columbia University, New York, 1932a.

[33] Thorndike, E. L.: *Selected Writings from a Connectionist Psychology.* Teachers College, Columbia University, New York, 1932b.

[34] Turing, A. M.: Intelligent Machinery. In Meltzer, B., Michie, D. (eds.), *Machine Intelligence*, volume 5, pages 3-23. Edinburgh University Press, Edinburgh, 1969.

[35] Wermter, S. and Sun, R (eds.): *Hybrid Neural Systems*. Springer Verlag, Berlin, 2000.