

# Hopfieldove neurónové siete a ich základné vlastnosti

Svorad Štolc Jr. (svoro@pobox.sk)  
MFF UK, Bratislava, Slovenská Republika

Marec 2000

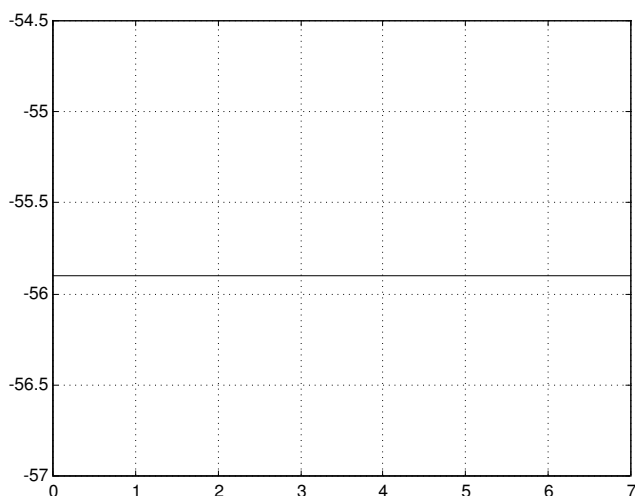
Hopfieldove siete sú jednou z klasických konštrukcií v oblasti umelých neurónových sietí. Ich pomocou je možné umelo modelovať množstvo zaujímavých biologiko-fyziologických fenoménov, z ktorých spomeniem predovšetkým auto-asociatívnu pamäť, generovanie alebo rozpoznávanie časových sekvencií, taktiež fenomén chýbajúcej základnej frekvencie a mnohé ďalšie. Rád by som sa v tejto práci zamerlal špeciálne na auto-asociatívnu pamäť a skúmanie jej parametrov, ktoré ovplyvňujú rozoznávanie zašumených vzorov.

Hopfieldova sieť použitá v tejto práci je skonštruovaná podľa klasickej schémy, jej dynamika je asynchrónna stochastická. K jej implementácii bol použitý jazyk integrovaný programe MatLab 5.0, nebol využitý žiaden hotový Toolbox, ani špeciálne funkcie zamerané na tvorbu neurónových sietí. Hlavným dôvodom pre výber tohoto vývojového prostredia bola jeho maticová orientácia, ako aj jednoduchosť tvorby grafických výstupov.

## 1 Teplota vs. minimalizácia energie

Teplota vystupujúca v hopfieldovom modeli predstavuje akúsi mieru deterministickosti (resp. stochastickosti) správania sa neurónov. V tejto časti by som rád objasnil akým spôsobom vplýva teplota na stabilitu systému v zmysle zotrvania v konkrétnom lokálnom minime energie. Sieť je zložená zo 100 neurónov (štvorec 10x10) a je natrénovaná na konkrétnych 5 vzorov ("U", "V", "X", "Y", "Z") očíslovaných v rovnakom poradí. Na vstup siete privádzame jeden z tréningových vzorov v neporušenej forme<sup>1</sup>, čím

<sup>1</sup>Ukázalo sa, že výber vzoru významne neovplyvní zistené výsledky. V tejto časti práce bol preto použitý vzor č. 1 ("U").

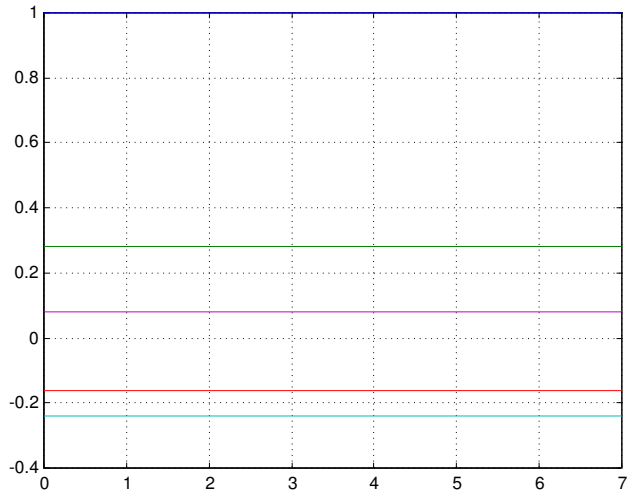


Obr. č. 1:  $T = 0$  (Vývoj energie systému.)

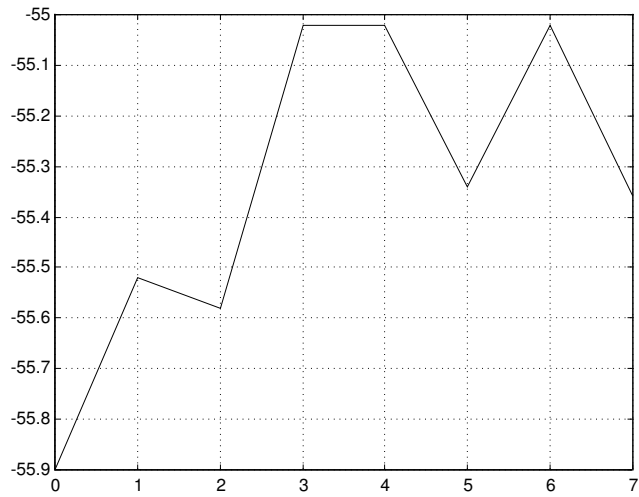
dosiahneme stav, ktorý je lokálnym minimom energie. Simulácia prebieha počas 7 epoch, nasledujúce grafy zachytávajú vývoj energie a prekryvov s tréningovými vzormi<sup>2</sup>. Simulácia bola opakovaná pri teplotách  $T = 0, 0.2, 0.4$  a  $1.0$  (viz. obr. č. 1 až 8).

Z grafov vyplýva, že teplota siete a variabilnosť prechodu z konkrétnej konfigurácie na jej nasledovníka spolu veľmi úzko súvisia. Ak má sieť nulovú teplotu, jej dynamika je deterministická, t.j. jej stav je stabilný počas všetkých krokov výpočtu (podobne aj prekryvy aktuálneho stavu s príslušnými vzormi korešpondujú s touto predstavou) (viz. obr. č. 1 a 2). Ak však teplotu zvyšujeme, systém začína oscilovať, nezostáva naďalej vo svojej počiatočnej stabilnej konfigurácii a žiaden z prekryvov už nie je konštantný. No aj pri teplote  $T = 0.4$  systém stále spoľahlivo udržuje dominantný prekryv so štartovacím vzorom. Pri zvýšení teploty na  $T = 1.0$  energia systému miestami stagnuje, celkový trend je však stúpajúci (viz. obr. č. 7). Predošlé úvahy môžeme zhrnúť do nasledujúceho tvrdenia: veľkosť teploty určuje schopnosť siete dostať sa z príslušne hlbokých miním energie.

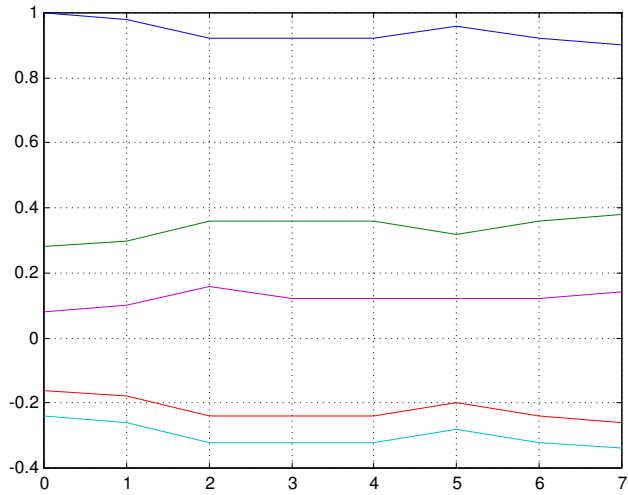
<sup>2</sup>Farebné označenie pre jednotlivé vzory: modrá ("U"); zelená ("V"); červená ("X"); tyrkysová ("Y"); fialová ("Z"). Toto označenie je v tejto práci implicitné.



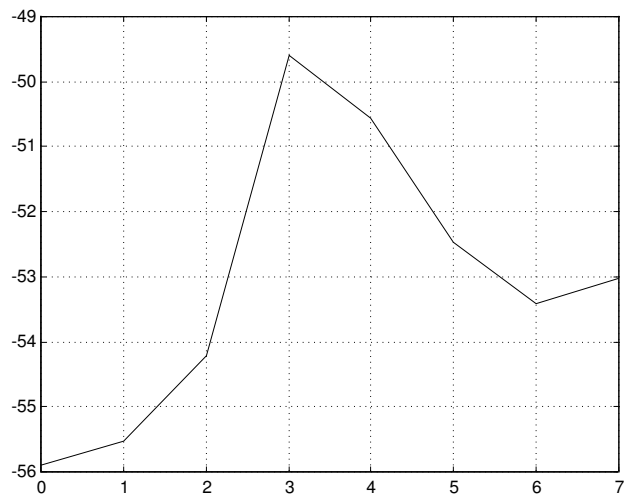
Obr. č. 2:  $T = 0$  (Vývoj překryvov s trénovacími vzormi.)



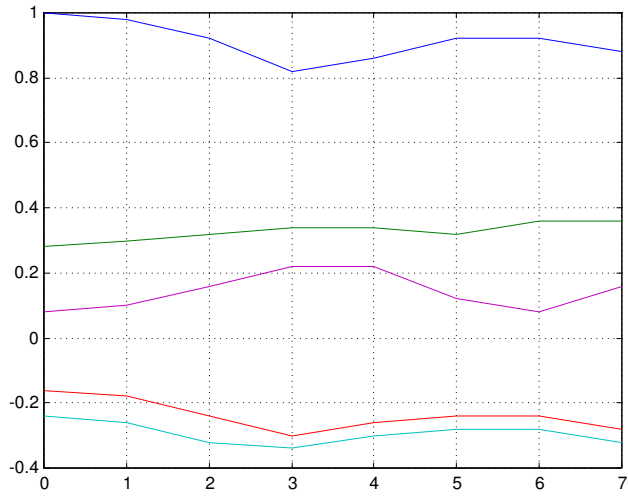
Obr. č. 3:  $T = 0.2$  (Vývoj energie systému.)



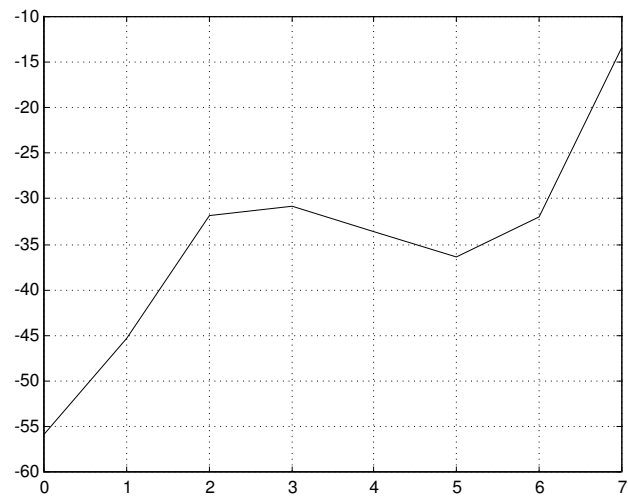
Obr. č. 4:  $T = 0.2$  (Vývoj překryvov s trénovacími vzormi.)



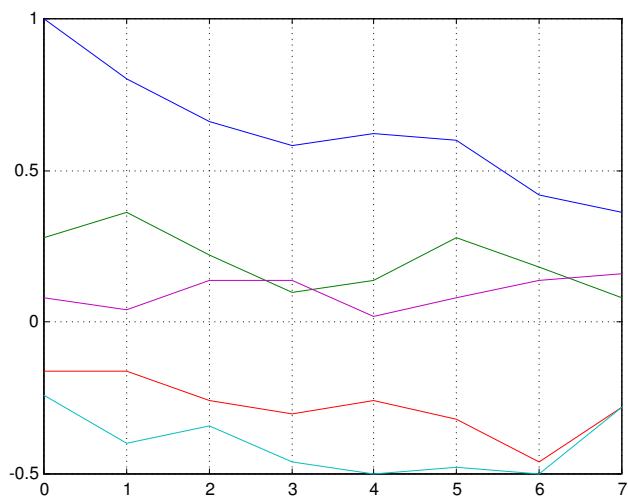
Obr. č. 5:  $T = 0.4$  (Vývoj energie systému.)



Obr. č. 6:  $T = 0.4$  (Vývoj překryvov s trénovacími vzormi.)



Obr. č. 7:  $T = 1$  (Vývoj energie systému.)



Obr. č. 8:  $T = 1$  (Vývoj prekryvov s tréningovými vzormi.)

Ak si všimneme oscilácie energie v uvedených prípadoch, zistíme, že rastú sigmoidálnym spôsobom (viz. obr. č. 9). Skúsme sa pozrieť na tento jav viac analyticky. Funkcia pravdepodobnosti

$$P(S_m = +1) = \frac{1}{1 + e^{\frac{2h_m}{T}}}$$

prechádza (z  $-\infty$ ) od nulových hodnôt k jednotkovým hodnotám (smerom k  $\infty$ ). Jej deriváciou podľa  $h_m$ <sup>3</sup> dostávame vzťah

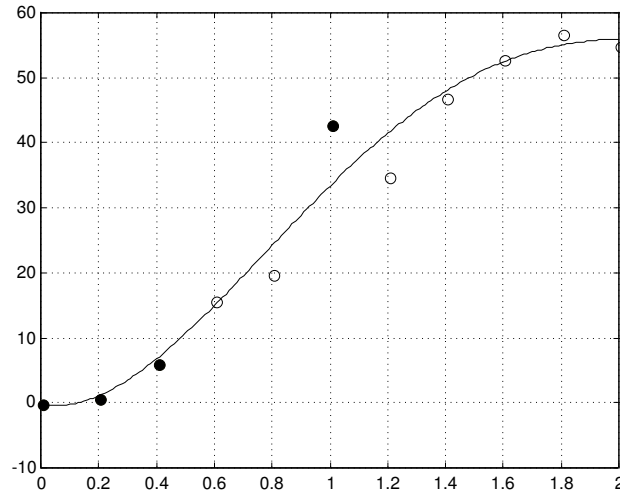
$$P'(h) = \frac{\partial P}{\partial h} = \frac{\frac{2e^{\frac{2h}{T}}}{T}}{\left(1 + e^{\frac{2h}{T}}\right)^2},$$

ktorého funkčná hodnota v bode 0<sup>4</sup> je rovná

$$P'(0) = \frac{1}{2T}.$$

<sup>3</sup> $h_m$  predstavuje EPSP pre  $m$ -tý neurón. V nasledujúcom vzťahu sa pre jednoduchosť požíva skrátené označenie  $h$ .

<sup>4</sup>Bod 0 je inflexným bodom a taktiež bodom jej najväčšieho stúpania pravdepodobnostnej funkcie  $P$ .



Obr. č. 9: Závislosť oscilácie energie (Y) od teploty systému (X). (Neoznačené hodnoty vychádzajú z nameraných údajov, ktoré nie sú spomínané v tejto práci.)

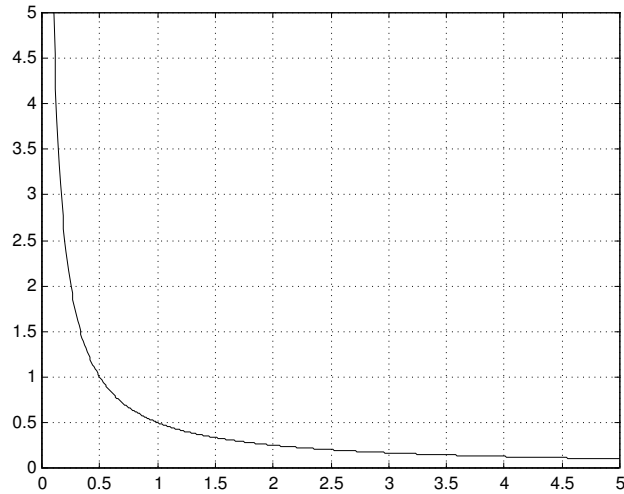
Čím viac sa hodnota posledného vzťahu blíži k nule, tým viac je pravdepodobnosť  $P(S_m = +1)$  podobná konštantnej 50%-nej pravdepodobnosti, ktorej derivácia je nulová na  $\mathfrak{R}$  (viz. obr. č. 10). Hodnoty prislúchajúce teplotám  $T \in (0, 0.2)$  sú dostatočne veľké a teda dynamika Hopfieldovej siete je "kvázi" deterministická<sup>5</sup>. Naopak pre teploty  $T \in (2, \infty)$ <sup>6</sup> je sklon pravdepodobnosti taký mierny, že sieť pracuje "kvázi" náhodne.

## 2 Auto-asociatívnosť a poškodené vzory

Ak by sme chceli použiť Hopfieldovu neurónovú sieť ako nástroj na rozpoznávanie vzorov, iste by nás zaujímali jej obmedzenia, hraničné prípady a taktiež jej optimálne nastavenia. Model HNS popisovaný v tejto práci

<sup>5</sup>Ak je  $P'(0)$  veľká, znamená to, že samotná pravdepodobnosť je v bode 0 prudká, čiže viac aproximuje funkciu  $S(x) = \begin{cases} 1 & \leftarrow x > 0 \\ 0 & \leftarrow x < 0 \end{cases}$ .

<sup>6</sup>Funkčné hodnoty pre  $T \in (0, 0.2)$  a  $T \in (2, \infty)$  sú si navzájom symetrické podľa osy hyperboly.



Obr. č. 10: Závislosť  $P'(0)$  (t.j. maximálneho stúpania  $P(S_m = +1)$ )(Y) od teploty systému (X).

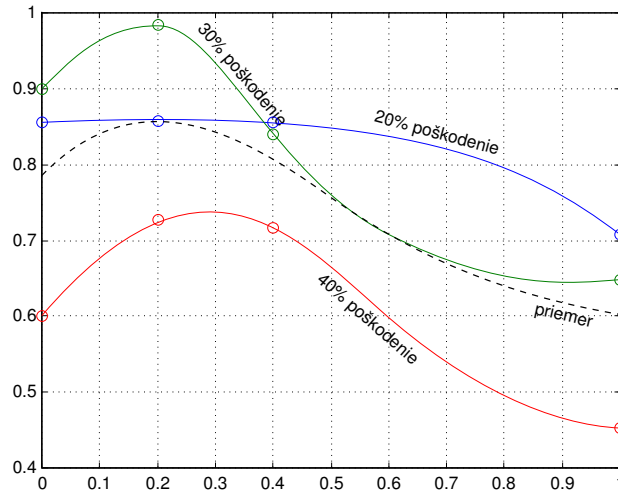
obsahuje jedinú konfiguračnú premennú (teplota), pomocou ktorej možno plynulo modifikovať úroveň stochastickosti systému a tým aj jeho celkovú funkčnosť. Pokúsil som sa preto odpovedať na dve základné otázky súvisiace s týmto problémom:

1. Aká je optimálna teplota systému, resp. pri akej teplote je systém najmenej omylný.
2. Ako veľmi môže byť poškodený vstupný vzor, aby ho systém s istou pravdepodobnosťou zaklasifikoval správne.

Odpovede možno založiť na štatistike získanej z nasledujúceho experimentu. Vyberieme jeden z tréningových vzorov<sup>7</sup> a predkladáme ho sieti v rôzne poškodenom stave (20%, 30% a 40%). Schopnosť rozoznávať daný vzor testujeme pre teploty  $T = 0, 0.2, 0.4$  a  $1$ , simuláciu opakujeme 5 krát pre každú kombináciu teploty a šumu, čím získame 5-ticu hodnôt, ktoré budú

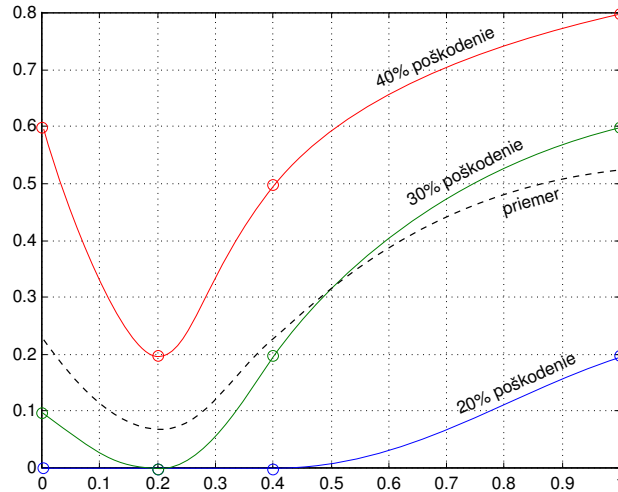
<sup>7</sup>V našom prípade bol použitý vzor č. 4 ("Y"), ktorý sa zdá byť najnáchyľnejší spadať do rôznych lokálnych miním. Frekvencovaným bolo minimum ekvivalentné lineárnej kombinácii "U" a "X".





Obr. č. 11: Závislosť prekryvu (Y) od teploty systému (X) a poškodenia počiatocného vzoru.

reprezentované v grafoch svojou priemernou hodnotou. Všímať si budeme dve veci. Za prvé maximálny prekryv s testovaným vzorom, táto vlastnosť nám pomôže zodpovedať prvú z nastolených otázok. Za druhé maximálny prekryv počas celého výpočtu, ktorý je kritériom klasifikácie testovaného vzoru, jej správnosti alebo nesprávnosti. Všimnime si obrázok č. 11, ktorý zachytáva prvú zo spomínaných vlastností. Možno si všimnúť, že maximá kriviek, odpovedajúcich rôznym stupňom zašumenia vstupu, sú sústredené v okolí  $T = 0.2$ . Ak teda chceme maximalizovať prekryvy so vstupným vzorom, najlepšie výsledky získame pre  $T \in \langle 0.1, 0.3 \rangle$  a to bez ohľadu na zašumenie vstupu. Podobne je to aj s chybou klasifikácie (vid. obr. č. 12), ktorej minimum je na tom istom mieste. Ak zvolíme teplotu zo spomínaného intervalu, s viac ako 70% pravdepodobnosťou vieme správne rekognoskovať vstupný vzor, ktorého poškodenie môže byť až 40%. Posledné tvrdenie je tiež opoveďou na druhú z našich otázok.



Obr. č. 12: Závislosť zlej klasifikácie (%) (Y) od teploty systému (X) a poškodenia počiatočného vzoru.

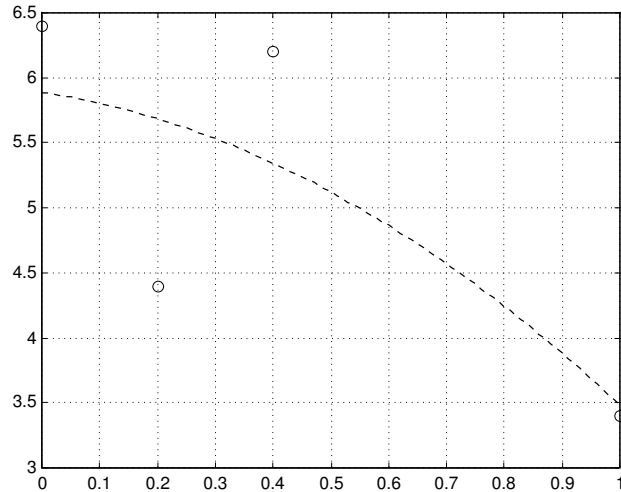
### 3 Klasifikácia náhodného vstupu

V predošlej časti sme sa zaoberali problémom klasifikácie poškodených vzorov. Skúsme sa teraz sústrediť na podobný problém, privádzajme na vstup HNS náhodný vzor<sup>8</sup>, ktorý je ekvivalentný 50%-ne poškodenému vzoru. Ak je metóda poškodzovania založená na pravdepodobnostne podmienenej negácii konkrétnej bunky, jej podobnosť s originálom sa pohybuje v intervale  $(-1, 1)$ . Ak zvolíme 10%-nú chybu, prekryv s originálom je približne 0.8. Ak je chyba 50%, prekryv je 0, z čoho vyplýva minimálna podobnosť (napr. 100% chyba  $\Rightarrow$  prekryv = -1).

Signál, ktorý privádzame sieti na vstup nie je odvodený od žiadneho tréningového vzoru, teda jeho prekryv s nimi je pre všetky približne rovnaký a zvyčajne dosť malý. Takýto stav implikuje vysokú vnútornú energiu, čiže nestabilitu. S klesajúcou energiou priemerná rozseparovanosť<sup>9</sup> prekryvov stúpa, čo možno pekne vidieť na obrázkoch č. 7 a 8. Ak je rozseparovanosť

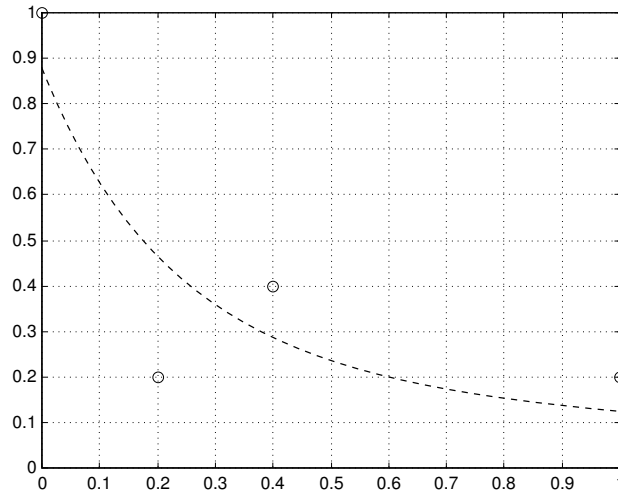
<sup>8</sup>1 a -1 sú vyberané s pravdepodobnosťou 50%.

<sup>9</sup>Priemerná rozseparovanosť prekryvov je  $\frac{\theta_{max} - \theta_{min}}{N}$ , kde  $\theta$  sú prekryvy a N je počet tréningových vzorov.



Obr. č. 13: Závislosť pomeru oscilácie prekryvov na začiatku a konci simulácie (Y) od teploty systému (X).

malá, konfigurácia siete je približne rovnako podobná všetkým vzorom, nezamieta, ani nepreferuje žiaden z nich. Zaujímavý je preto pomer finálnej a úvodnej rozseparovanosti. Čím je táto väčšia, tým úspešnejšia je relaxácia siete, tým zreteľnejšie sú rozdiely medzi prekryvmi a tým ľahšie vieme určiť "vítězný" vzor. Ak je však pomer malý, sieť stagnuje buď v nejakom lokálnom minime, alebo sa pohybuje cez spolu nesúvisiace náhodné konfigurácie (tzv. chaotická trajektória). Závislosť spomínaného pomeru na teplote systému (viz. obr. č. 13) má klesajúci priebeh, na intervale  $\langle 0, 1 \rangle$  sa zníži hodnota takmer na polovicu, hoci k najväčšiemu poklesu dochádza medzi 0.4 a 1. Samotný fakt, že systém preferuje jeden konkrétny vzor a to aj v prípade náhodného vstupu, nám nepomôže pri hádaní "vítěza" už v nulte pozícii. Môže existovať viacero spôsobov, ktoré pomôžu tento problém riešiť, napríklad veľmi jednoduchým pravidlom môže byť výber vzoru s najväčším štartovacím prekryvom. O tom, či je táto heuristika adekvátne nám hovorí obrázok č. 14. V podstate sa ňou možno riadiť iba pri veľmi nízkych teplotách (cca.  $T \in \langle 0, 0.1 \rangle$ ).

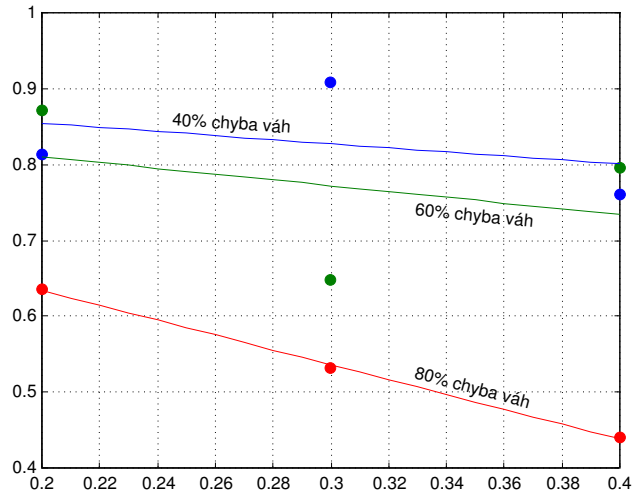


Obr. č. 14: Závislosť preferovania úvodného maximálneho prekryvu (%) (Y) od teploty systému (X).

## 4 Odolnosť voči poškodeniu váhovej matice

Ďalšou dôležitou vlastnosťou, ktorá sa zvykne testovať v súvislosti s neurónovými sieťami, je robustnosť alebo odolnosť voči náhodnému zanikaniu synaptických spojení. Robustnosť je vlastná biologickým neurónovým sieťam, napr. v istých oblastiach nervovej sústavy človeka je možné distribuovane ničiť veľké množstvá synaptických spojení (napr. pomocou alkoholu), no na funkčnosti celku sa zdanlivo nič nemení... V našom prípade bola robustnosť testovaná pre rôzne poškodený vzor č. 4 ("Y") a konštantné  $T = 0.2$ <sup>10</sup>. Zaujímavá je závislosť maximálneho prekryvu so vzorom "Y", od zašumenia vstupu (vid. obr. č. 15). Podobný (inverzný) vývoj má aj závislosť nesprávnej klasifikácie od šumu (vid. obr. č. 16). Z oboch grafov je zreteľne vidno, že zmena vlastností siete nie je v lineárnom vzťahu k poškodeniu váhovej matice. Rozdiel medzi stavmi s  $40\%W_{err}$  a  $60\%W_{err}$  je podstatne menší ako medzi  $60\%W_{err}$  a  $80\%W_{err}$ , z čoho vyplýva, že aj pri 60%-nom zániku synaptických spojení spoľahlivosť siete neklesne pod 70%, čo je naozaj uspokojujúce pri tak mohutnom zásahu

<sup>10</sup>Optimálnosti  $T = 0.2$  je venovaná časť č. 2 tejto práce.

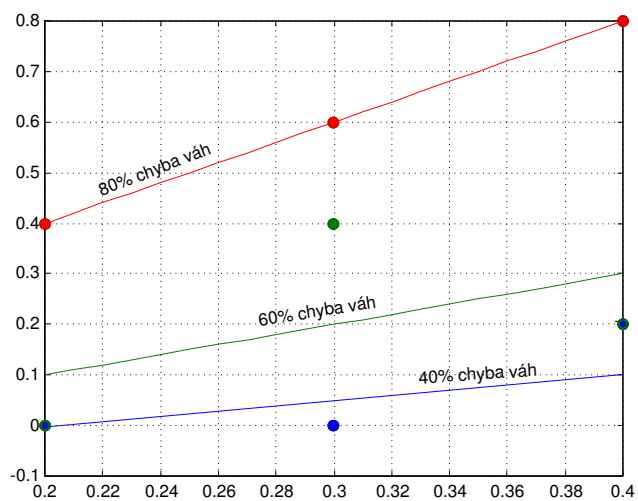


Obr. č. 15: Závislosť maximálneho prekryvu s úvodným vzorom (Y) od jeho zašumenia (X) a poškodenia váhovej matice.

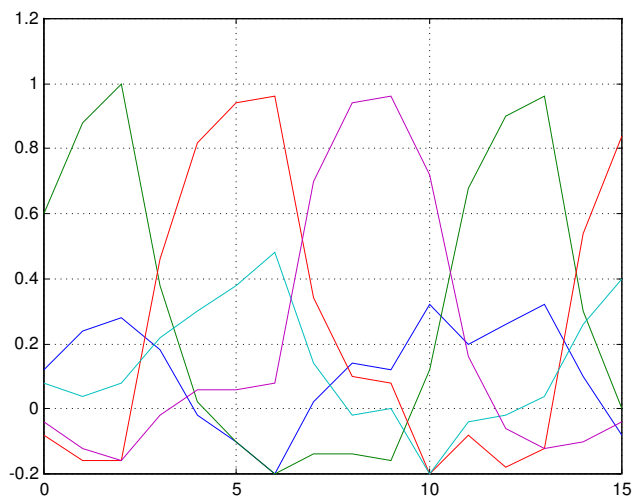
do štruktúry siete.

## 5 Generovanie časových postupností

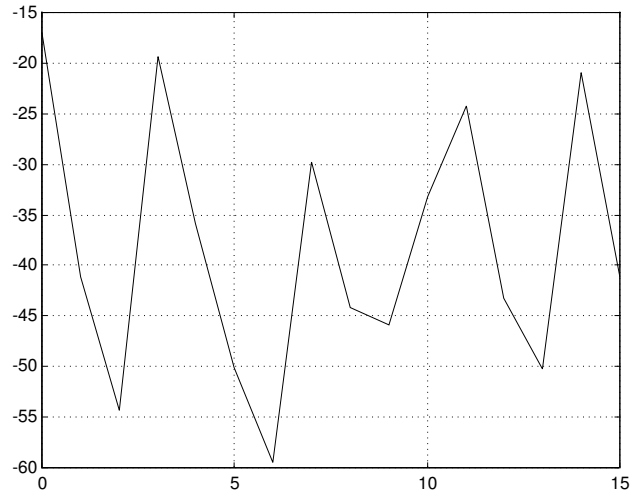
Schopnosť HNS generovať prípadne rozpoznávať časové sekvencie vzorov je spôsobená takzvanými pomalými synaptickými spojeniami. Tieto istou mierou ovplyvňujú veľkosť postsynaptického potenciálu a to vždy s presne určeným oneskorením (perioda striedania vzorov). Sieť možno natrénovať na rôzne sekvencie, podstatné je, že sa nemôžu prekryvať. Príklad generatívnej činnosti HNS je zaznamenaný obrázkami č. 17 až 19. Celkom typický je vlnitý vývoj prekryvov a energie, sieť pravidelne strieda nízko a vysoko energetické stavy, ktoré odzrkadľujú prechody z jednej spádovej oblasti do druhej. Naša sieť bola natrénovaná na postupnosť "V", "X", "Z", "V", čo zapríčinilo cyklickosť generovanej sekvencie. Parametre simulácie boli  $T = 0.2$ ,  $\tau = 3$  a  $\lambda = 2$ .



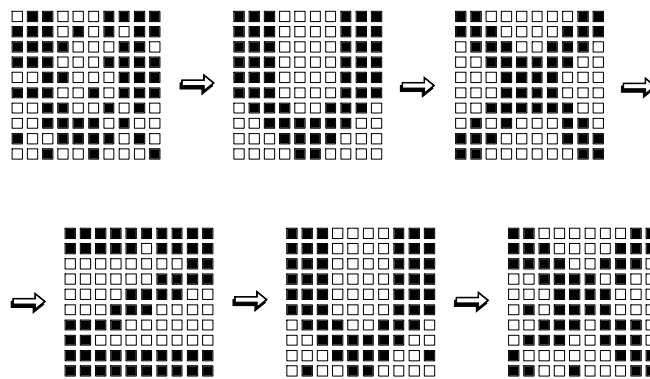
Obr. č. 16: Závislosť nesprávnej klasifikácie úvodného vzoru (%) (Y) od jeho zašumenia (X) a poškodenia váhovej matice.



Obr. č. 17: Vývoj prekryvov s tréningovými vzormi počas generovania časovej postupnosti.



Obr. č. 18: Vývoj energie systému počas generovania časovej postupnosti.



Obr. č. 19: Stavy siete korešpondujúce s 0., 1., 6., 9., 13. a 15. krokom simulácie (viď. obr. č. 17 a 18).

## 6 Záver

Hopfieldove siete sú veľmi jednoduchou abstrakciou skutočných neurónových sietí, no paradoxne už tu možno sledovať fenomény, ktoré obyčajne spájame s vyššou nervovou činnosťou. Ich skúmanie na nízkej úrovni nám môže významne pomôcť preniknúť do povahy istých mozgových procesov. HNS nám odkrývajú zaujímavý náhľad na schopnosť asociovať, reagovať alebo rozpoznávať. Metóda, pomocou ktorej riešia spomenuté problémy je zaujímavá predovšetkým svojou jednoduchosťou, transparentnosťou a univerzálnosťou. Tá istá sieť je rovnako dobrým klasifikátorom ako generátorom, vlastnosti platiace pre jednu aplikáciu platia podobne pre druhú. Pozoruhodná je tiež robustnosť použitej štruktúry, sieť pracuje bez väčších zmien aj s doslova zničenou váhovou maticou. Nevýhodou je pomerne vysoká výpočtová zložitosť (približne  $O(n^2)$ ), mohutnosť siete 10x10 neurónov bola za daných podmienok na hranici únosnosti.