Neuronske mreže: Rekurzivne mreže

Prof. dr. sc. Sven Lončarić

Doc. dr. sc. Tomislav Petković

Fakultet elektrotehnike i računarstva

https://www.fer.unizg.hr/predmet/neumre_c

Pregled predavanja

- Uvod
- Hopfieldova mreža
    - Diskretna Hopfieldova mreža
    - Faze spremanja i dohvata
    - Energetska (Ljapunovljeva) funkcija
    - Lažna stanja
    - Memorijski kapacitet
- Stohastički strojevi
    - Boltzmannov stroj
- Simulirano hlađenje

Uvod

- Višeslojne i radijalne neuronske mreže su dobro poznati primjeri mreža bez povratnih veza ili unaprijednih mreža
  - veze između neurona ne stvaraju petlje
- U ovom predavanju razmatramo rekurzivne i/ili povratne mreže
  - veze između neurona stvaraju usmjerene petlje
- Razmotrit ćemo i stohastičke strojeve koji koriste povratne veze
  - neuroni su stohastički

Uvod

- Glavne značajke rekurzivnih neuronskih mreža:
  - veze između neurona tvore usmjerene petlje
  - usmjerene petlje sadrže neki oblik vremenskog kašnjenja odnosno radi se o *povratnoj vezi*, a ne *algebarskoj petlji*
  - pokazuju dinamičko vremensko ponašanje
- Modeliranje višestrukih povratnih veza je teško
  - mreža može postati nestabilna
- Iskoristive arhitekture uglavnom ograničavaju broj povratnih veza ili uvode druga ograničenja kako bi analiza mreže bila moguća

Kratka povijest

- Rekurzivne neuronske mreže
  - Hopfieldova neuronska mreža (*Hopfield 1982, Little 1974*)
  - Dvosmjerna asocijativna memorija (eng. BAM; *Kosko 1988*)
  - Duga kratkoročna memorija (eng. LSTM, od *long short-term memory; Hochreiter & Schmidhuber 1997*)
- Stohastički strojevi
  - Boltzmannov stroj (*Hinton, Terrence & Ackley 1984*)
  - Ograničen Boltzmannov stroj (*Smolensky 1986*)
  - Sigmoidalne mreže (*Neal 1992*)
  - Helmholtzov stroj (*Dayan 1995, Hinton 1995*)

Hopfieldova neuronska mreža

- Hopfieldova neuronska mreža sastoji se od skupa neurona i od pripadnih kašnjenja koji zajedno tvore sustav višestrukih povratnih veza
  - izlaz svakog neurona se dovodi na ulaze svih ostalih neurona
  - nema vlastitih povratnih veza (neuron prema samom sebi)
- Takva arhitektura pokazuje nova svojstva koja nisu očita (eng. *emergence*)
  - sadržajem adresirana memorija (eng. CAM, od *content addressable memory*)
- Arhitektura je jednostavna i omogućuje punu analizu
  - razmotriti ćemo stabilnost mreže i kapacitet memorije

Hopfieldova mreža kao sadržajem adresirana memorija

- Svaki uzorak je zapamćen u nekom lokalnom minimumu energetske funkcije
  - zapamćeni uzorak zovemo fundamentalnom memorijom
  - to je atraktor energetske funkcije
  - svaki atraktor odgovara najnižoj točki neke doline energetske funkcije
- Kroz rad mreže njeno stanje evoluira iz nekog početnog stanja u najbliži atraktor
  - sva početna stanja koja teže istom atraktoru čine bazen konvergencije (eng. *basin of attraction*)

Diskretna Hopfieldova mreža

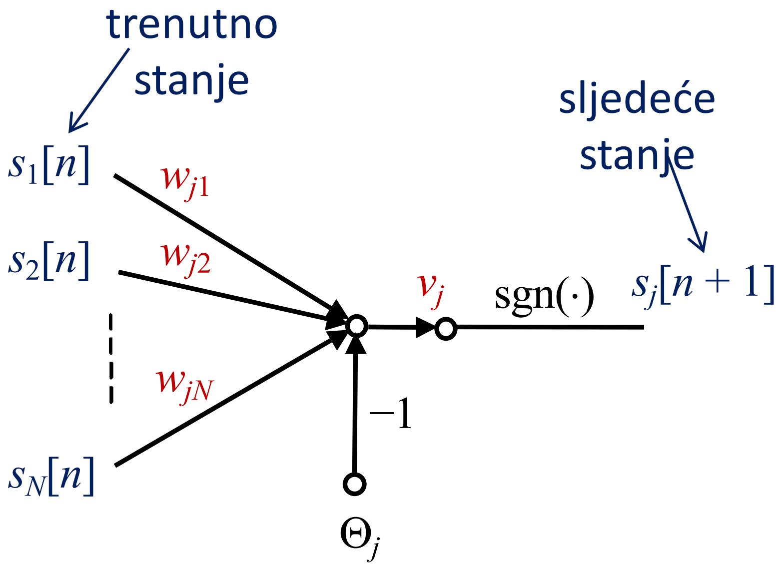
- Diskretna Hopfieldova mreža se temelji na McCulloch-Pittsovom modelu neurona
  - izlaz svakog neurona je ili +1 ili -1
- Stanje diskretne Hopfieldove mreže koja se sastoji od N neurona je vektor stanja $\mathbf{s} = [s_1, s_2,\ldots, s_N]^T$, gdje je s_j izlaz iz j-tog neurona
- Veza između i-tog i j-tog neurona ima težinu w_{ji}
  - w_{ji} određuje doprinos stanja s_i u aktivaciji j-tog neurona
  - nema vlastitih povratnih veza, odnosno vrijedi $w_{ii} = 0$

Diskretna Hopfieldova mreža

- Aktivacijski potencijal v_j neurona j jest:

$$v_j = \sum_{i=1}^{N} w_{ji} s_i - \Theta_j$$

$$s_j[n+1] = \mathrm{sgn}(v_j)$$



Nema vlastitih povratnih veza, tj. $w_{jj} = 0$

Diskretna Hopfieldova mreža

- Izlaz *j*-tog neurona jest

$$s_j = \begin{cases} +1, & v_j > 0 \\ -1, & v_j < 0 \end{cases}$$

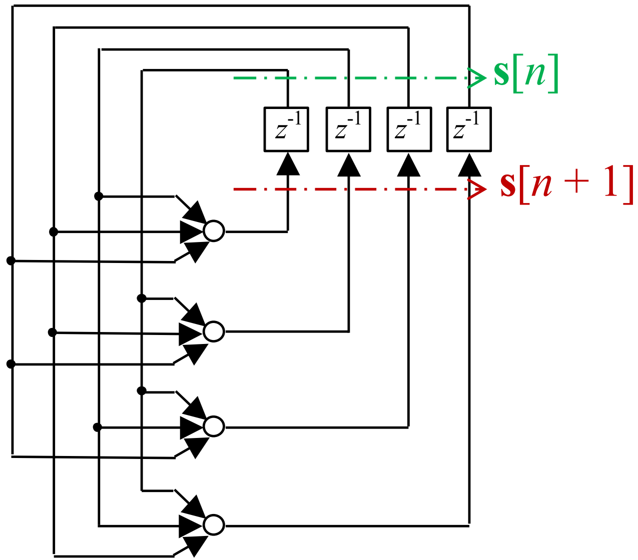
što za $v_j \neq 0$ postaje

$$s_j = \operatorname{sgn}(v_j)$$

- Ako je $v_j = 0$ onda *j*-ti neuron ostaje u svom sadašnjem stanju, odnosno izlaz se ne mijenja
  - to je samo jedan od više mogućih izbora
  - takav odabir međutim čini dijagram toka *simetričnim*

Diskretna Hopfieldova mreža

- Arhitektura diskretne Hopfieldove mreže sastavljene od četiri neurona ($N = 4$)
- Blokovi označeni sa z^{-1} su elementi za kašnjenje
  - izlaz iz elementa za kašnjenje jest trenutno stanje mreže
  - ulaz u element za kašnjenje jest sljedeće stanje mreže



Diskretna Hopfieldova mreža

- Postoje dvije različite faze rada Hopfieldove mreže:
  - faza pamćenja (ili učenja)
  - faza dohvata (ili eksploatacije)
- U fazi pamćenja mreža pamti zadane podatke
- U fazi dohvata neki vektor se postavlja kao početno stanje mreže; mreža zatim dohvaća najsličniji spremljeni uzorak nakon što stanje mreže konvergira k najbližem stabilnom stanju
  - dohvat spremljenih podataka nije trenutan

Diskretna Hopfieldova mreža: Faza pamćenja

- Pretpostavimo da želimo pohraniti skup od p N-dimenzionalnih vektora:

$$\{\xi_m \mid m = 1, \ldots, p\}$$

- Tih p vektora zovemo fundamentalne memorije
- Neka je $\xi_{m,i}$ *i*-ti element vektora ξ_m s vrijednošću ili +1 ili -1
- Prema pravilu vanjskog produkta (*generalizacija Hebbovog pravila učenja*) sinaptička težina koja povezuje *i*-ti s *j*-tim neuronom jest:

$$w_{ji} = \frac{1}{N} \sum_{m=1}^{p} \xi_{mj} \xi_{mi}$$

Diskretna Hopfieldova mreža: Faza pamćenja

- Za diskretnu Hopfieldovu mrežu vrijedi $w_{ii} = 0$ za svaki i
  - Nema vlastitih povratnih veza
- Neka je $\mathbf{W}$ *matrica sinaptičkih težina* dimenzija $N \times N$
- Matricu $\mathbf{W}$ možemo izračunati prema izrazu (slijedi iz izraza za težinu i iz svojstva $w_{ii} = 0$):

$$\mathbf{W} = \frac{1}{N} \sum_{m=1}^{p} \xi_m \xi_m^{\ T} - \frac{p}{N} \mathbf{I}$$

vektorski ili vanjski umnožak fundamentalne memorije ξ_m

Diskretna Hopfieldova mreža: Faza pamćenja

- Iz izraza za računanje *matrice sinaptičkih težina* $\mathbf{W}$ ponovo uočavamo sljedeće:
  - Izlaz svakog neurona se dovodi na ulaze svih preostalih neurona
  - Nema vlastite povratne veze neurona u samog sebe, odnosno $w_{ii} = 0$ za svaki i ($\mathbf{W}$ ima nule na glavnoj dijagonali)
  - Matrica sinaptičkih težina $\mathbf{W}$ je simetrična, odnosno $w_{ji} = w_{ij}$

Diskretna Hopfieldova mreža: Faza dohvata

- U fazi dohvata N-dimenzionalni vektor $\mathbf{x}$ čiji elementi su ili $+1$ ili -1 se postavlja kao početno stanje diskretne Hopfieldove mreže
  - vektor $\mathbf{x}$ se zove ispitni uzorak (eng. *probe*)
- Ispitni vektor $\mathbf{x}$ uobičajeno predstavlja nepotpunu ili zašumljenu verziju fundamentalne memorije mreže

Diskretna Hopfieldova mreža: Faza dohvata

- Nakon što se ispitni uzorak $\mathbf{x}$ postavi kao stanje mreže ona dalje radi koristeći pravilo *asinkronog* (ili *serijskog*) osvježavanja:

  Svaki neuron iz mreže nasumično, no ipak s nekom jednolikom učestalosti, računa svoj aktivacijski potencijal te osvježava svoj izlaz ako je potrebno.

- Slijedi:
  - Odabir neurona koji se osvježava je nasumičan
  - Promjena stanja je deterministička
- Opisana procedura osvježavanja se ponavlja sve dok se stanje mijenja

Diskretna Hopfieldova mreža: Faza dohvata

- Kada više nema promjena stanja mreže kažemo da je mreža dosegla vremenski stalan vektor stanja $\mathbf{s}$ koji ispunjava uvjet stabilnosti:

$$s_j = \mathrm{sgn}\left(\sum_{i=1}^{N} w_{ji} s_i - \Theta_j \right), \quad j = 1, 2, \dots, N,$$

koji u matričnom zapisu postaje

$$\mathbf{s} = \mathrm{sgn}(\mathbf{W}\mathbf{s} - \boldsymbol{\Theta})$$

gdje je $\mathbf{W}$ matrica sinaptičkih težina i gdje je $\boldsymbol{\Theta}$ vektor pragova (ponekad se pomak $\mathbf{b} = -\boldsymbol{\Theta}$ koristi umjesto pragova $\boldsymbol{\Theta}$).

18

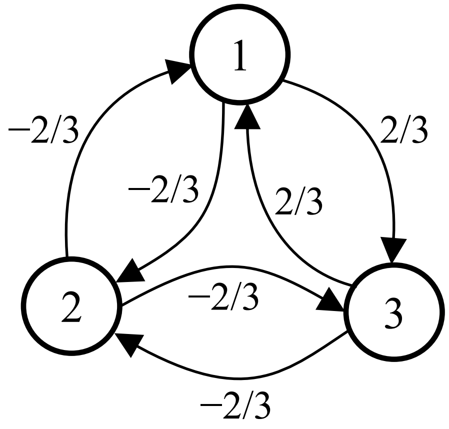
Diskretna Hopfieldova mreža: Faza dohvata

- Stanje mreže $\mathbf{s}$ koje ispunjava uvjet stabilnosti se naziva *stabilnim stanjem* ili *fiksnom točkom u* prostoru stanja sustava
  - može se označiti i s y jer se radi o izlazu mreže
- Diskretna Hopfieldova mreža uvijek konvergira u stabilno stanje ako se dohvat vrši *asinkrono* (ili *serijski*)
- Alternativni pristup je pravilo *sinkronog* (ili *paralelnog*) osvježavanja za koje se izlazno stanje istovremeno određuje za sve neurone
  - tada stanje konvergira ili u fiksnu točku ili u granični ciklus maksimalne duljine 2

Diskretna Hopfieldova mreža: Primjer 14.2 iz Haykina

- Zadana je Hopfieldova mreža sastavljena od tri neurona koja ima matricu težina $\mathbf{W}$
- Matrica $\mathbf{W}$ je simetrična i ima nule na glavnoj dijagonali
- Neka je vektor pragova nula, dakle $\Theta = 0$

$$\mathbf{W} = \frac{1}{3}\begin{bmatrix} 0 & -2 & +2 \\ -2 & 0 & -2 \\ +2 & -2 & 0 \end{bmatrix}$$



Diskretna Hopfieldova mreža: Primjer 14.2 iz Haykina

- Za tri neurona postoji sveukupno $2^3 = 8$ mogućih stanja mreže
- Od tih osam mogućih stanja samo dva su stabilna
- Stabilna stanja zadovoljavaju uvjet stabilnosti

$$\mathbf{y} = \operatorname{sgn}(\mathbf{W}\mathbf{y} - \boldsymbol{\Theta})$$

- Stabilna stanja ili fiksne točke su:
  - $[+1, -1, +1]^T$
  - $[-1, +1, -1]^T$

Diskretna Hopfieldova mreža: Primjer 14.2 iz Haykina

- Provjerimo zadovoljava li stanje $y = [+1, -1, +1]^{\mathrm{T}}$ uvjet stabilnosti:

$$\mathbf{Wy} = \frac{1}{3}\begin{bmatrix} 0 & -2 & +2 \\ -2 & 0 & -2 \\ +2 & -2 & 0 \end{bmatrix}\begin{bmatrix} +1 \\ -1 \\ +1 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} +4 \\ -4 \\ +4 \end{bmatrix}$$

$$\operatorname{sgn}(\mathbf{Wy}) = \operatorname{sgn}\left(\frac{1}{3}\begin{bmatrix} +4 \\ -4 \\ +4 \end{bmatrix}\right) = \begin{bmatrix} +1 \\ -1 \\ +1 \end{bmatrix} = \mathbf{y}$$

Diskretna Hopfieldova mreža: Primjer 14.2 iz Haykina

- Preostalih šest stanja su nestabilna što možemo jednostavno provjeriti jer ne zadovoljavaju uvjet stabilnosti
- Mreža ima dva stabilna stanja:
  - $[+1, -1, +1]^{\mathrm{T}}$
  - $[-1, +1, -1]^{\mathrm{T}}$
- Da su dva stabilna stanja zaista dvije ($p = 2$) fundamentalne memorije možemo provjeriti i računanjem matrice sinaptičkih težina:

$$\mathbf{W} = \frac{1}{N}\sum_{m=1}^{p} \xi_m \xi_m^{\ T} - \frac{p}{N}\mathbf{I}$$

Diskretna Hopfieldova mreža: Primjer 14.2 iz Haykina

- Matrica sinaptičkih težina jest:

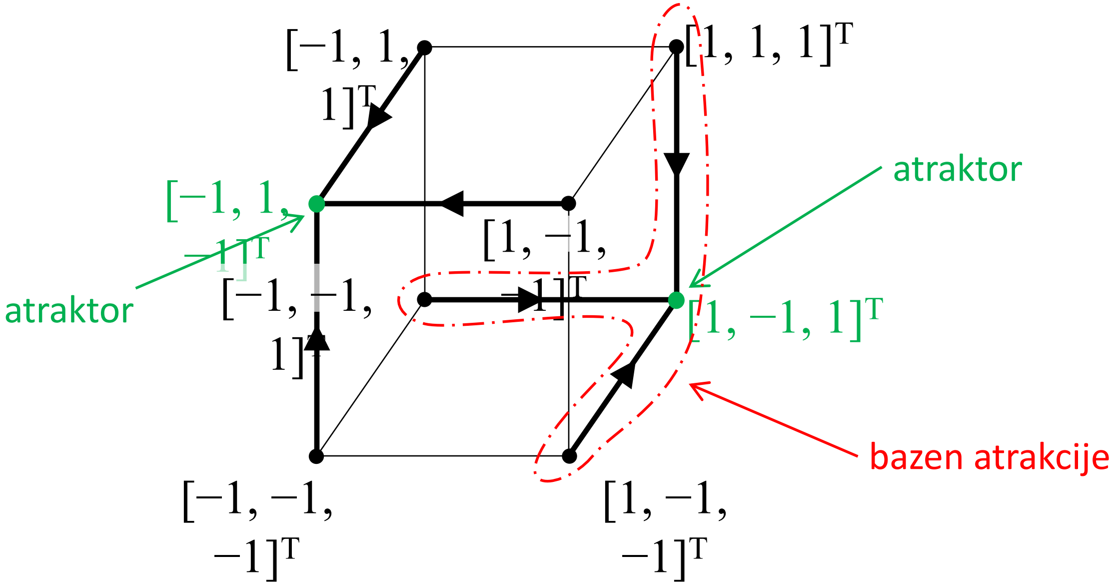
$$\mathbf{W} = \frac{1}{3}\begin{bmatrix} +1 \\ -1 \\ +1 \end{bmatrix}\begin{bmatrix} +1 & -1 & +1 \end{bmatrix} + \frac{1}{3}\begin{bmatrix} -1 \\ +1 \\ -1 \end{bmatrix}\begin{bmatrix} -1 & +1 & -1 \end{bmatrix} - \frac{2}{3}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \frac{1}{3}\begin{bmatrix} 0 & -2 & +2 \\ -2 & 0 & -2 \\ +2 & -2 & 0 \end{bmatrix}$$

- Uočite da je dobivena matrica sinaptičkih težina **W** jednaka početnoj matrici zadanoj u primjeru.

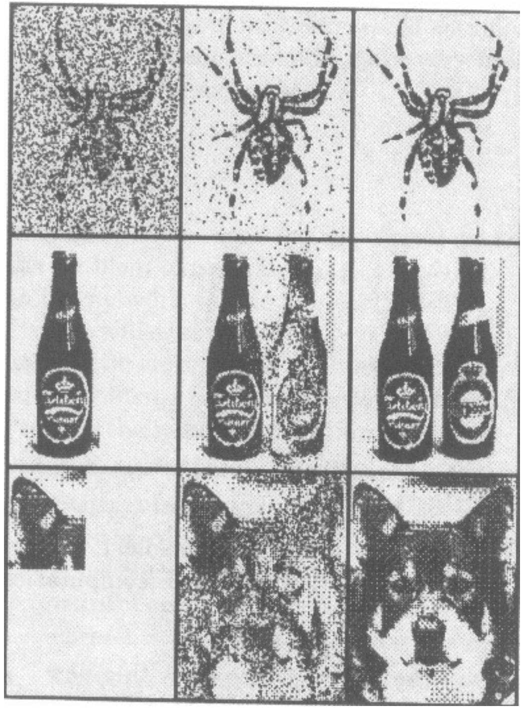
Diskretna Hopfieldova mreža: Primjer 14.2 iz Haykina

- Prostor stanja Hopfieldove mreže:
  - zelene točke označavaju stabilna stanja (fiksne točke, atraktori)
  - strelice pokazuju smjer promjene prema stabilnim stanjima



Primjene Hopfieldove mreže

- Upotreba Hopfieldove mreže kao asocijativne slikovne memorije koja može rekonstruirati slike
  - Slika 2.1 iz knjige Hertz, John A., Anders S. Krogh, i Richard G. Palmer. “*Introduction to the theory of neural computation*”. 1991.
  - korištena je rijetko povezana Hopfieldova mreža koja pamti sedam binarnih slika veličine 130×180 piksela
  - srednji stupac prikazuje međustanje tijekom faze dohvata



Diskretna Hopfieldova mreža: Energetska funkcija

- Energetska funkcija diskretne Hopfieldove mreže kad je $\Theta = 0$ jest:

$$E = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N} w_{ji} s_i s_j = -\tfrac{1}{2}\mathbf{s}^T \mathbf{W}\mathbf{s}$$

- Dodavanjem konstante izrazu E (tako da je $E \geq 0$) energetska funkcija postaje Ljapunovljeva funkcija Hopfieldove mreže
  - korisna za analizu stabilnosti mreže
- Za fiksne težine w_{ji} energija ovisi o vektoru stanja mreže $\mathbf{s} = [s_1, s_2, \ldots, s_N]^T$

Energetska funkcija se uvijek smanjuje s vremenom

- Pretpostavimo da je u procesu asinkronog osvježavanja kod faze dohvata j-ti neuron promijenio svoje stanje s_j ($+1$ u -1 ili -1 u $+1$)
- Promjena ukupne energije mreže uzrokovana tom promjenom stanja j-tog neurona je:

$$\Delta E = -\Delta s_j \sum_{i=1, i \neq j}^{N} w_{ji} s_i$$

- Pažljivom analizom gornjeg izraza za promjenu energije ΔE primjećujemo da se ukupna energija mreže E uvijek smanjuje sve dok se ne dostigne lokalni minimum energetske funkcije
  - Pokažimo to na sljedeće dvije prikaznice

Energetska funkcija se uvijek smanjuje s vremenom

Neka je desni član u produktu ΔE pozitivan:

$$\sum_{i=1, i \neq j}^{N} w_{ji} s_i > 0$$

Tada je novo stanje s_j za j-ti neuron također pozitivno:

$$s_j = \operatorname{sgn}\left(\sum_{i=1, i \neq j}^{N} w_{ji} s_i \right) = +1, \quad \Delta s_j = 2$$

Kako su oba člana produkta u ΔE pozitivna ukupna promjena mora biti negativna zbog negativnog predznaka:

$$\Delta E = -\Delta s_j \sum_{i=1, i \neq j}^{N} w_{ji} s_i < 0$$

Energetska funkcija se uvijek smanjuje s vremenom

Neka je desni član u produktu ΔE negativan:

$$\sum_{i=1, i\neq j}^{N} w_{ji} s_i < 0$$

Tada je novo stanje s_j za j-ti neuron također negativno:

$$s_j = \mathrm{sgn}\left(\sum_{i=1, i\neq j}^{N} w_{ji} s_i \right) = -1, \quad \Delta s_j = -2$$

Kako su oba člana produkta u ΔE negativna njihov umnožak je pozitivan pa je ukupna promjena opet negativna zbog predznaka:

$$\Delta E = -\Delta s_j \sum_{i=1, i\neq j}^{N} w_{ji} s_i < 0$$

Energetska funkcija se uvijek smanjuje s vremenom

- Pokazali smo da je ΔE uvijek negativan
- To znači da je energetska funkcija E diskretne Hopfieldove mreže padajuća funkcija s vremenom
  - Vrijedi samo za asinkrono osvježavanje stanja mreže
- Energija se prestaje smanjivati kada mreža dosegne stabilno stanje
  - Stabilna stanja ili fiksne točke odgovaraju minimumima energetske funkcije

Diskretna Hopfieldova mreža: Energetska (Ljapunovljeva) funkcija

- Postojanje Ljapunovljeve funkcije implicira stabilnost:

Stavak

Neka je $s = s_0$ ravnotežna točka diferencijske jednadžbe

$$s(n+1) = f\left(s(n)\right)$$

gdje je $f: D \to \mathbb{R}^N$ lokalno Lipschitzeva u $D \subset \mathbb{R}^N$ i gdje je $s_0 \in D$. Neka postoji funkcija $E: D \to \mathbb{R}$ koja je kontinuirana i koja je takva da vrijedi

$$E(0) = 0 \text{ i } E(s - s_0) > 0,\ \forall\ s \in D \backslash \{s_0\}$$

$$E(f(s - s_0)) - E(s - s_0) \leq 0,\ \forall\ s \in D$$

Tada je ravnotežna točka $s = s_0$ stabilna.

Hopfieldova mreža kao asocijativna memorija

- Za fiksne težine w_{ji} energija ovisi o vektoru stanja mreže $\mathbf{s} = [s_1, s_2, \ldots, s_N]^T$
- Ovisnost energije o stanju mreže može se prikazati kao energetski krajolik s zamišljenim brdima i dolinama
- Lokalni minimumi energetskog krajolika su smješteni u dolinama i ponašaju se kao atraktori
- Dva uvjeta moraju biti ispunjena kako bi se Hopfieldova mreža ponašala kao asocijativna memorija su:
  - Zapamćeni uzorci moraju biti atraktori
  - Svaki zapamćeni uzorak mora imati bazen atrakcije dovoljne veličine

Hopfieldova mreža kao asocijativna memorija: Lažna stanja

- Ako pokušavamo zapamtiti veliki broj fundamentalnih memorija (uzoraka) ξ_m korištenjem Hebbovog pravila onda se pojavljuju lažna stanja (eng. *spurious states*)
- Lažna stanja s_{spurious} su stanja koja se NE nalaze u skupu fundamentalnih memorija ξ_m, $m = 1,\ldots, p$, ali koja odgovaraju minimumima energetske funkcije
- Razlikujemo tri tipa lažnih stanja:
  1. Zrcalna stanja
  2. Miješana stanja
  3. Stanja tipa spinskog stakla

Hopfieldova mreža kao asocijativna memorija: Lažna zrcalna stanja

- Energetska funkcija Hopfieldove mreže je simetrična
  - Promjena predznaka svih aktivacijskih potencijala v_j rezultira istom energijom
  - Ako energetska funkcija postiže lokalni minimum za fundamentalnu memoriju ξ onda lokalni minimum postoji i za $-\xi$
  - $-\xi$ je zrcalno ili obrnuto stanje
- Prema tome zrcalna refleksija fundamentalne memorije se ponaša kao fundamentalna memorija
  - Problem može biti ublažen u praksi: ako zrcalno stanje nije fundamentalna memorija možemo uspostaviti konvenciju za pretvorbu između ξ i $-\xi$

Hopfieldova mreža kao asocijativna memorija: Lažna mješovita stanja

- Zbrojevi ili razlike neparnog broja fundamentalnih memorija mogu također biti lokalni minimumi energetske funkcije
- Primjer je mješavina triju stanja $\mathbf{s}_{\text{mix}} = \text{sgn}(\xi_1 + \xi_2 + \xi_3)$
  - U prosjeku $\mathbf{s}_{\text{mix}}$ odgovara predznaku stanja ξ_i 6 od 8 puta
  - Dakle, $\mathbb{E}[\sum_j \xi_{i,j}\, \mathbf{s}_{\text{mix},j}] = 3N/4$ za $i = 1, 2, 3$
  - Ako je preslušavanje malo (vidi prikaznicu br. 40) tada vrijedi $\mathbf{s}_{\text{mix},j} = \text{sgn}(\sum_i w_{ji}\, \mathbf{s}_{\text{mix},j})$
- Vjerojatnost pojavljivanja lažnih stanja raste s p
- Vlastita povratna veza generira lažne atraktore

Hopfieldova mreža kao asocijativna memorija: Lažna stanja tipa spinskog stakla

- Ova stanja nisu povezana s bilo kojom konačnom kombinacijom fundamentalnih memorija
- Naziv dolazi od bliske veze s *modelom spinskog stakla* u *statističkoj mehanici (fizici)*
- Nastaju ako je broj uzoraka p koje želimo zapamtiti prevelik
  - Povećanjem p matrica sinaptičkih težina $\mathbf{W}$ počinje nalikovati slučajnoj matrici
  - Član u zbroju $\mathrm{sgn}(\sum_i w_{ji}\,\mathbf{x}_{\mathrm{probe},i})$ koji određuje neku fundamentalnu memoriju je nadjačan od strane svih preostalih članova
  - Associjativna memorija počinje izgledati kao *spinsko staklo*

Kapacitet pohrane Hopfieldove mreže

- Fundamentalne memorije Hopfieldove mreže nisu uvijek stabilna stanja
- Mogu se pojaviti lažna stanja koja odgovaraju stabilnim stanjima mreže koja su različita od fundamentalnih memorija
- Ove dvije pojave smanjuju efikasnost Hopfieldove mreže kao asocijativne memorije
- Korištenjem vjerojatnosne analize moguće je ispitati stabilnost fundamentalnih memorija
    - Primijetite da analiza koju ćemo provesti zanemaruje lažna stanja
- Na temelju takve analize može se procijeniti kapacitet pohrane Hopfieldove mreže

Kapacitet pohrane Hopfieldove mreže

- Razmotrimo stabilnost j-tog neurona za fundamentalnu memoriju $\xi_n = [\xi_{n,1}, \xi_{n,2}, \ldots, \xi_{n,N}]^T$
  - j-ti neuron se često naziva i j-tim bitom
- Uvjet stabilnosti jest:

$$\xi_{nj} = \mathrm{sgn}(v_{nj})$$

$$v_{nj} = \sum_{i=1}^{N} w_{ji}\xi_{ni}$$

pri čemu su težine w_{ji} dane izrazom

$$w_{ji} = \frac{1}{N}\sum_{m=1}^{p}\xi_{mj}\xi_{mi}$$

Kapacitet pohrane Hopfieldove mreže

- Aktivacijski potencijal j-tog neurona za fundamentalnu memoriju ξ_n je:

$$v_j = \sum_{i=1}^{N}\left(\frac{1}{N}\sum_{m=1}^{p}\xi_{m,j}\xi_{m,i}\right)\xi_{n,i} = \frac{1}{N}\sum_{m=1}^{p}\xi_{m,j}\sum_{i=1}^{N}\xi_{m,i}\xi_{n,i}$$

- Izdvajanjem člana $\xi_{n,j}$ iz sume dobivamo:

signal

preslušavanje (šum)

$$v_{nj} = \xi_{nj} + \frac{1}{N}\sum_{i=1}^{N}\sum_{m=1,m\neq n}^{p}\xi_{mj}\xi_{mi}\xi_{ni}$$

- Ako je drugi pribrojnik (preslušavanje) jednak nuli onda je aktivacijski potencijal j-tog neurona v_j stabilan jer je zadovoljen uvjet stabilnosti:

$$\xi_{nj} = \mathrm{sgn}(v_{nj})$$

Kapacitet pohrane Hopfieldove mreže

- Zbog aktivacijske funkcije $\mathrm{sgn}(\cdot)$ izlaz iz j-tog neurona je stabilan čak i u slučaju slabog preslušavanja
  - slabo preslušavanje ne može promijeniti predznak aktivacijskog potencijala v_j
- Elementi svake fundamentalne memorije ξ_n su ili $+1$ ili -1
  - onda su težine w_{ji} u intervalu $[+1, -1]$
- Prema tome ako je broj fundamentalnih memorija p dovoljno malen ($p << N$) onda je preslušavanje zanemarivo

Kapacitet pohrane Hopfieldove mreže

- Neka su elementi svake fundamentalne memorije ξ_n Bernoulijeve slučajne varijable na skupu $\{-1, +1\}$
  - Svaki element u ξ_n ima srednju vrijednost 0 i varijancu $\sigma^2 = 1/N^2$
- Preslušavanje je sada zbroj $N(p - 1)$ međusobno nezavisnih jednako distribuiranih slučajnih varijabli
- Prema *centralnom graničnom teoremu* razdioba zbroja nezavisnih jednako distribuiranih slučajnih varijabli konvergira po distribuciji u normalnu ili Gaussovu razdiobu
- Prema tome preslušavanje slijedi Gaussovu razdiobu sa srednjom vrijednosti 0 i varijancom $N(p - 1) \cdot 1/N^2 = (p - 1)/N$

Kapacitet pohrane Hopfieldove mreže

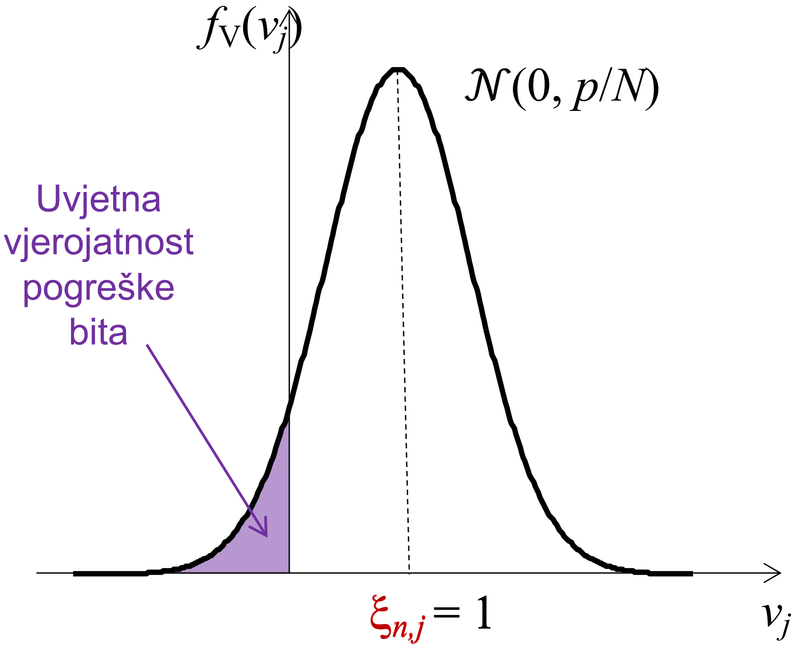
- Sada možemo izračunati odnos signal-šum ili SNR:

$$\rho = \frac{\text{signal}}{\text{šum}} = \frac{1}{(p-1)/N} = \frac{N}{p-1} \cong \frac{N}{p}$$

- Zapamćene komponente fundamentalne memorije su stabilne samo ako je odnos signal šum velik
  - prema tome mora biti $N >> p$
- Inverz SNR-a se naziva faktor ispune α
  - Smanjenje faktora ispune $\alpha = p/N$ umanjuje vjerojatnost pogreške

Kapacitet pohrane Hopfieldove mreže

- Postoji kritična vrijednost faktora ispune α_c nakon koje se učinak mreže značajno smanjuje
  - $\alpha_c = 0.138$
- Tada je kapacitet mreže $p_{max} = 0.138\ N$ što daje 0.36% nestabilnih izlaznih bitova
  - To je kapacitet mreže s pogreškama



Kapacitet pohrane Hopfieldove mreže

- Možemo tražiti da pri radu mreže uvjetna vjerojatnost pogreške bude manja od $1/N$
- U tom slučaju maksimalni kapacitet jest $p_{\max} \approx N/(2 \log_e N)$
    - To je kapacitet pohrane *gotovo bez pogrešaka*
- Naposlijetku, možemo tražiti da se sve fundamentalne memorije ispravno dohvaćaju
- Takva je maksimalni kapacitet $p_{\max} \approx N/(4 \log_e N)$
- Kapaciteti pohrane navedeni ovdje su izvedeni iz asimptotskog ponašanja uvjetne vjerojatnosti pogreške bita

Stohastičke mreže

- Stohastičke mreže uvode slučajne (ili stohastičke) varijacije u mrežu:
  - stohastičke aktivacijske funkcije
  - stohastičke sinaptičke težine
- Prilagođene su za rješavanje optimizacijskih zadataka
  - problemi pretraživanja kao primjerice pretraga uz zadovoljenje ograničenja (CSP, eng. *constraint-satisfaction problem*)
  - problemi učenja kao primjerice generiranje podataka
- Obično takve mreže imaju pridruženu energetsku funkciju koja se minimizira
  - želimo naći globalni minimum umjesto lokalnog minimuma
  - stohastička priroda mreže nam omogućava izbjegavanje lokalnih minimuma

Boltzmannov stroj

- Naziv dolazi od Boltzmannove razdiobe koja se koristi u fizici (statistička mehanika)
- Sastoji se od mnogo identičnih neurona koji su potpuno povezani jedan s drugim
  - svaki neuron ima dva stanja, uključen (on) ili isključen (off)
  - stanje je vjerojatnosna funkcija stanja svih ostalih neurona i ovisi samo o sinaptičkim težinama, tj. aktivacijska funkcija je stohastička
  - sve veze su dvosmjerne i simetrične
- Moguće je odabrati aktivacijsku funkciju koja djeluje na ponašanje individualnih neurona tako da se minimizira globalna energija

Boltzmanov stroj i Hopfieldova mreža

- Boltzmanov stroj se može promatrati kao stohastička Hopfieldova mreža
  - sinaptičke težine su determinističke
  - aktivacijske funkcije su stohastičke
- Međutim, svrha mreže je drugačija
  - Hopfieldova mreža pamti zadane uzorke u lokalnim minimumima svoje energetske funkcije, prema tome Hopfieldova mreža traži lokalne minimume
  - Boltzmanov stroj mora pak pronaći globalni minimum svoje energetske funkcije

Boltzmannov stroj: Stohastička aktivacijska funkcija

- Stohastička aktivacijska funkcija neurona u Boltzmanovom stroju je inspirirana Metropolisovim algoritmom:
  - Usko je povezana s metodom simuliranog hlađenja (žarenja)
- Stohastička aktivacijska funkcija i-tog neurona je

$$s_{i,\text{next}} = \begin{cases} 1, & \text{s vjerojatnošću } p = \dfrac{1}{1+e^{-\Delta E_i/T}} \\ -1, & \text{s vjerojatnošću } p = \dfrac{e^{-\Delta E_i/T}}{1+e^{-\Delta E_i/T}} \end{cases}$$

gdje je

$$\Delta E = -2 s_{i,\text{previous}} \sum_{j=1}^{N} w_{ij} s_{j,\text{previous}}$$

promjena energije zbog promjene stanja i-tog neurona

- T je temperatura

Boltzmannov stroj: Stohastička aktivacijska funkcija

- Kao i kod Hopfieldove mreže neuron koji se osvježava se odabire proizvoljno (asinkrono osvježavanje)
- Svojstvo aktivacijske funkcije jest da se neuron može postaviti i u lokalno ne-optimalno stanje s nekom vjerojatnošću
  - drugim riječima *energija mreže ponekad raste*
  - vjerojatnost odabira ne-optimalnog stanja se smanjuje s padom temperature
  - na temperaturi $T = 0$ aktivacijska funkcija postaje deterministička i Boltzmanov stroj se pretvara u Hopfieldovu mrežu
- U slučaju kada se temperatura T ne mijenja mreža nakon mnogo promjena stanja eventualno dostiže termalnu ravnotežu
- Odnos vjerojatnosti dva energetska stanja A i B mreže koja je u termalnoj ravnoteži slijedi Boltzmanovu razdiobu: $$\frac{P(A)}{P(B)} = e^{-(E_A - E_B)/T}$$

Boltzmannov stroj: Tijek ili raspored hlađenja

- Na niskim temperaturama postoji značajna pristranost prema stanjima niske energije, no vrijeme potrebno da se dostigne termalna ravnoteža može biti jako dugo
- Na visokim temperaturama termalna ravnoteža se dostiže brzo, no nema pristranost prema stanjima niske energije
- Prema tome možemo započeti s visokom temperaturom koju onda postupno snižavamo (kontrolirano hlađenje ili žarenje)
  - Na visokim temperaturama time vršimo grubu pretragu u prostoru svih globalnih stanja sustava
  - Snižavanjem temperature nalazimo bolje minimume polazeći od grubog minimuma odabranog na prethodnoj višoj temperaturi
- Brzina snižavanje temperature definira se tijekom ili rasporedom hlađenja

Boltzmannov stroj: Pitanje konvergencije

- Hopfieldova mreža je deterministička i uvijek konvergira k najbližem lokalnom minimumu
- Boltzmanov stroj je stohastički i konvergira k globalnom minimumu u očekivanju
  - Nema jamstva da je pronađeni minimum zaista i globalni minimum
  - Možemo samo procijeniti vjerojatnost tog ishoda
  - Vjerojatnost je tim veća što je tijek hlađenja (žarenja) sporiji i što je vrijeme simulacije dulje
- U praksi se stroj uobičajeno pojednostavljuje kroz ograničenja na povezanost neurona
  - Ograničen Boltzmanov stroj (*Smolensky 1986*)

Neuronske mreže: Boltzmannov stroj

SIMULIRANO HLAĐENJE

Simulirano hlađenje

- Predloženo 1983. godine od Kirkpatricka et al.
- Simulirano hlađenje ili simulirano žarenje (eng. *simulated annealing*) je probabilistički algoritam za aproksimaciju globalnog optimuma neke funkcije
  - Smatra se *meta-heurističkim* postupkom
  - Ima čvrste temelje u *termodinamici* i *statističkoj mehanici*
- Simulirano hlađenje traži globalne ekstreme *ograničene funkcije cilja* oko određenog *konfiguracijskog područja*
  - Može se koristiti za rješavanje mnogih kombinatornih optimizacijskih problema

Simulirano hlađenje

- Osnovni koncept i naziv dolaze iz postupka žarenja u metalurgiji
  - metal se zagrije i onda se polako hladi u kontroliranim uvjetima sa svrhom dobivanja poželjnih svojstava
  - kontrolirano polagano hlađenje se zove žarenje (eng. *annealing*)
  - potiče difuziju čime metal napreduje prema svom stanju termalne ravnoteže
  - brzo hlađenje (kaljenje, eng. *quenching*) onemogućava fazni prijelaz i metal ostaje u metastabilnom stanju
  - prema tome kod pravilnog žarenja metal postiže stanje minimalne energije
- Postupak se zove i simulirano žarenje što je izravni prijevod engleskog naziva *simulated annealing*
  - Napomena: kaljenje (eng. *quenching*) je naglo hlađenje tako da ponekad korišteni naziv „simulirano kaljenje” nije prikladan

Simulirano hlađenje: Metropolisov algoritam

- Metropolis i suradnici su 1953. opisali kako simulirati skupinu molekula u termalnoj ravnoteži na nekoj stalnoj temperaturi $T\,[K]$
- U simulaciji se nasumično odabrana molekula slučajno pomiče nakon čega se računa promjena energije ΔE [J] cijele skupine
- Glavni rezultat je Metropolisov kriterij prihvaćanja koji definira vjerojatnost $\Pr$ za prihvaćanje simulirane promjene energije

$$\Pr\{\text{prihvati } \Delta E\} = \begin{cases} 1, & \Delta E \leq 0 \\ e^{-\Delta E / kT}, & \Delta E > 0 \end{cases}$$

- Drugim riječima
  - Ako promjena smanjuje ukupnu energiju sustava onda se prihvaća
  - Ako promjena povećava ukupnu energiju sustava onda se prihvaća s vjerojatnošću $\exp(-\Delta E/kT)$, gdje je k Boltzmannova konstanta

Simulirano hlađenje: Metropolisov algoritam

- Ako se simulacija izvodi za dovoljan broj slučajnih pomicanja (koraka) tada je konačan razmještaj molekula blizak onome u termalnoj ravnoteži ili ustaljenom stanju
  - To je globalni minimum na temperaturi $T\,[K]$
- Formalni dokaz konvergencije modelira opisanu simulaciju kao homogeni *Markovljev lanac* za kojeg se pokazuje da njegovo stabilno stanje odgovara termalnoj ravnoteži
  - U teoriji se konvergencija postiže tek za beskonačan broj koraka simulacije

Simulirano hlađenje: Metropolisov algoritam

- Tri pretpostavke uz koje Metropolisov algoritam postiže termalnu ravnotežu su:

1) Reverzibilnost (simetrija)
   - vjerojatnost odabira sljedećeg stanja je jednaka kao i vjerojatnost povratka iz tog sljedećeg stanja u trenutno stanje
2) Ergodičnost
   - slučajni pomaci molekula su takvi da molekule mogu postići bilo koji položaj u svom konfiguracijskom prostoru
3) Konvergencija ka kanonskoj razdiobi
   - vjerojatnosti u kriteriju prihvaćanja su takve da u prosjeku ansambl teži u Boltzmannovu (ili Gibbsovu) razdiobu, odnosno

$$F(\text{konačno stanje}) \propto \exp\left(-\frac{E_{\min}}{kT}\right)$$

Simulirano hlađenje: Metropolisov algoritam

- Metropolisov algoritam rezultira termalnom ravnotežom ili ustaljenim stanjem na nekoj odabranoj temperaturi
- Proces simuliranog hlađenja iterativno izvodi Metropolisov algoritam (*Kirkpatrick 1983*):
  - Prvo sustav kojeg želimo optimirati rastalimo na nekoj visokoj temperaturi
  - Na jako visokim temperaturama sva energetska stanja su skoro pa jednako vjerojatna
  - Zatim polako spuštamo temperaturu sve dok se sustav ne zamrzne te nema više promjena
  - Na svakoj temperaturi Metropolisova simulacija se mora izvoditi dovoljno dugo kako bi sustav dostigao ustaljeno stanje na toj temperaturi

Simulirano hlađenje: Formalni opis problema

- Neka je Ω skup svih konfiguracija (područje pretraživanja)
- Neka je $f{:}\Omega \rightarrow \mathbb{R}$ funkcija cilja (ili energetska funkcija)
  - tj. vrijednost $f(\omega)$ je cijena nekog rješenja $\omega \in \Omega$
- Neka je $N(\omega)$ funkcija susjedstva za $\omega \in \Omega$
  - tj. $N(\omega)$ su susjedna rješenja koja se mogu dosegnuti unutar jedne iteracije lokalnog algoritma pretraživanja iz trenutnog stanja $\omega \in \Omega$
- Neka je T_n pseudo-temperatura sustava u iteraciji n
  - tj. Boltzmannova konstanta je uklopljena u T_n
- Želimo pronaći rješenje $\omega^* \in \Omega$ takvo da je

$$f(\omega^*) \leq f(\omega)$$

za sve $\omega \in \Omega$

Simulirano hlađenje: Unutarnja petlja na odabranoj temperaturi

- Započinjemo s nekim početnim rješenjem $\omega_{\mathrm{steady}} \in \Omega$ s prethodne više temperature T_{n-1}
  - Ako rješenje ω_{steady} nije dostupno tada proizvoljno odaberemo bilo koje rješenje
- Slučajno biramo susjedno rješenje $\omega' \in N(\omega)$
- Prihvaćamo odabrano rješenje ω' ako ono ispunjava Metropolisov kriterij prihvaćanja:

$$\Pr\{\text{prihvati } \omega'\} = \begin{cases} 1, & \text{ako } f(\omega') - f(\omega) \le 0 \\ e^{-\left(f(\omega') - f(\omega)\right)/T_n}, & \text{ako } f(\omega') - f(\omega) > 0 \end{cases}$$

- Ponavljamo opisani slučajni odabir rješenja sve dok se ne dostigne ustaljeno stanje ω_{steady} na temperaturi T_n
  - U praksi najčešće samo ponavljamo proceduru odabira M_n puta

Simulirano hlađenje: Vanjska petlja i tijek hlađenja

- Vanjska petlja jednostavno odabire sljedeću temperaturu T_{n+1} i ponavlja unutarnju petlju
  - T_n mora biti padajući niz vrijednosti po n koji teži u nulu

- Tijek ili raspored hlađenja definira:
  - Kroz koliko koraka M_n je potrebno izvoditi Metropolisov algoritam na svakoj temperaturi T_n
  - Koji padajući niz temperatura T_n koristimo

Pseudokod algoritma

Odaberite početno rješenje $\omega \in \Omega$

Odaberite padajući niz temperatura T_n

Odaberite broj koraka ponavljanja M_n za svaku temperaturu T_n

Postavite brojač ponavljanja vanjske petlje u $n = 0$

Ponavljajte

    Postavite brojač ponavljanja unutarnje petlje u $m = 0$

    Ponavljajte

        Odaberite slučajno rješenje $\omega' \in N(\omega)$

        Izračunajte $\Delta = f(\omega') - f(\omega)$

        Ako $\Delta \leq 0$ onda $\omega \leftarrow \omega'$

        Ako $\Delta > 0$ onda

            Odaberite slučajni broj ξ iz $\mathrm{U}[0,1]$ razdiobe

            Ako $\xi < \exp(-\Delta / T_n)$ onda $\omega \leftarrow \omega'$

        $m \leftarrow m + 1$

    Sve dok ne bude $m = M_n$

    $n \leftarrow n + 1$

Sve dok se ne ispuni kriterij zaustavljanja

Simulirano hlađenje: Konvergencija

- Konvergencija k globalnom minimumu je dokazana korištenjem teorije Markovljevih lanaca
  - U suštini svi dokazi pokazuju da se *globalni minimum* podudara sa *stabilnim stanjem* Markovljevog lanca koji modelira simulirani proces hlađenja
- Takva formalna analiza pokazuje da postoje ograničenja na proces hlađenja, točnije na padajući niz temperatura (*Mitra, 1986*)

$$T_n \geq \frac{\beta}{\log(n)}$$

gdje je β konstanta koja ovisi o problemu kojeg rješavamo

Simulirano hlađenje: Rasprava

- Ključno svojstvo algoritma simuliranog hlađenja jest mogućnost odabira rješenja koje je lokalno lošije
  - To svojstvo ga čini drugačijim od jednostavnih postupaka kao što su metoda uspona ili metoda gradijentnog spusta koji u svakoj iteraciji uvijek odabiru bolje rješenje
- Ovo ključno svojstvo čini algoritam „glupim” (Fleischer 1995):
  - *“…it is blind to the global optimum and can find it only to leave it in the next iteration searching the entire configuration space in total ignorance of the quality of the solution it left…” („…on je slijep za globalni optimum i može ga pronaći pa ga ostaviti odmah u sljedećoj iteraciji pretraživanja čitavog konfiguracijskog prostora s potpunim nepoznavanjem kvalitete rješenja kojega je odbacio…”)*

Simulirano hlađenje: Praktična razmatranja

- U praksi želimo zadržati svojstvo konvergencije uz kratko vrijeme izvršavanja
  - Radi se o oprečnim zahtjevima
- Tri glavna problema koja se trebaju uzeti u obzir kod praktične primjene simuliranog hlađenja su:
  1. Prikladan tijek (raspored) hlađenja (žarenja)
  2. Prikladna funkcija cilja
  3. Dobro definirana struktura susjedstva
- Nažalost, rješenja navedenih problema gotovo uvijek ovise o konkretnom zadatku kojeg rješavamo

Problem trgovačkog putnika

- Problem trgovačkog putnika jest:
  - *Ako je zadana lista gradova i cijene putovanja između svakog para gradova potrebno je pronaći najpovoljniji put kojim se svaki grad posjeti samo jednom i na čijem kraju se vraćamo u polazni grad.*
- Opisani problem je NP-težak
- Za $N \geq 3$ postoji točno $(N-1)!/2$ mogućih neusmjerenih putova
  - Prema tome složenost iscrpnog pretraživanja jest $\mathrm{O}(n!)$
  - Postojeći algoritmi koji nalaze egzaktno rješenje imaju složenost $\mathrm{O}(n^2 2^n)$
- *Simulirano hlađenje* može pronaći približno rješenje
  - U mnogim primjenama takvo sub-optimalno rješenje je sasvim zadovoljavajuće

Neuronske mreže: Rekurzivne mreže

ZADACI ZA VJEŽBU

Zadaci

Zadatak 14.4 iz (Haykin, 1998) – Hopfieldova mreža

Promatramo Hopfieldovu mrežu koja se sastoji od 5 neurona i koja mora zapamtiti sljedeće tri zadane fundamentalne memorije:
$\xi_1 = [1, 1, 1, 1, 1]^T$, $\xi_2 = [1, -1, -1, 1, -1]^T$ and $\xi_3 = [-1, 1, -1, 1, 1]^T$.
Neka je prag svih neurona jednak nuli.

(a) Izračunajte 5x5 matricu sinaptičkih težina mreže.

(b) Koristeći pravilo asinkronog osvježavanja pokažite da sve tri fundamentalne memorije ξ_1, ξ_2 i ξ_3 ispunjavanju uvjet stabilnosti.

(c) Istražite uspješnost dohvata kada se na ulaz mreže postavi zašumljena varijanta memorije ξ_1 u kojoj je promijenjen predznak drugog elementa, odnosno ispitni vektor jest $x_{\text{probe}} = [1, -1, 1, 1, 1]^T$.

Zadaci

Zadatak 14.7 iz (Haykin, 1998) – Hopfieldova mreža

Zadana je Hopfieldova mreža koja se sastoji od dva neurona i čija matrica sinaptičkih težina jest:

$$\mathbf{W} = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$$

Pragovi svih neurona su jednaki nuli.

Četiri moguća stanja mreže su

$\mathbf{s}_1 = [1,\ 1]^T$, $\mathbf{s}_2 = [-1,\ 1]^T$, $\mathbf{s}_3 = [-1,\ -1]^T$ i $\mathbf{s}_4 = [1,\ -1]^T$.

(a) Pokažite da su stanja $\mathbf{s}_2$ i $\mathbf{s}_4$ stabilna pomoću uvjeta stabilnosti.

(b) Pokažite da su stanja $\mathbf{s}_2$ i $\mathbf{s}_4$ stabilna pomoću energetske funkcije.

(c) Pokažite da stanja $\mathbf{s}_1$ i $\mathbf{s}_3$ čine granični ciklus. Koja je duljina tog graničnog ciklusa?

Zadaci

Boltzmannov stroj

Promatramo Boltzmannov stroj koji radi na temperaturi $T = 1$.

Neka energije dva stanja označena s A i B budu $E_A = -25$ i $E_B = -22$.

Osim toga znamo da se sustav između svih mogućih stanja nalazi u nekom od ta dva stanja, no ne znamo u kojem.

(1) Izračunajte $P(A)/P(B)$ (taj omjer se zove *Boltzmannov faktor*).

(2) Izračunajte uvjetnu vjerojatnost $P(A|A \text{ ili } B)$, odnosno izračunajte vjerojatnost da je stroj u stanju A ako znamo da se stroj nalazi u jednom od stanja A i B.

(3) Izrazite uvjetnu vjerojatnost $P(A|A \text{ ili } B)$ samo pomoću energije i temperature.

Literatura

1. Haykin, Simon. *“Neural Networks: A Comprehensive Foundation” (2nd Edition)*. Prentice Hall, 1998.
2. John A., Anders S. Krogh, and Richard G. Palmer. *“Introduction to the theory of neural computation”*. Westview Press. June 1991.
3. Hopfield, John J. *“Neurons with graded response have collective computational properties like those of two-state neurons.”* Proceedings of the National Academy of Sciences 81.10 (1984): 3088-3092.
4. Hinton, Geoffrey E., Terrence J. Sejnowski, and David H. Ackley. *“Boltzmann machines: Constraint satisfaction networks that learn” (Technical Report CMU-C5-84-119)*. Carnegie-Mellon University. 1984.
5. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. “*Equation of state calculations by fast computing machines*”. The journal of chemical physics, 21(6), 1087-1092. (1953)
6. Kirkpatrick, Scott, C. Daniel Gelatt, and Mario P. Vecchi. “*Optimization by simulated annealing*.” Science 220.4598 (1983): 671-680.
7. Henderson, Darrall, Sheldon H. Jacobson, and Alan W. Johnson. “*The theory and practice of simulated annealing*.” Handbook of metaheuristics. Springer US, 2003. 287-319.