Neuronske mreže
Algoritmi obučavanja

Milan M.Milosavljević

ADALINA (Adaptive Linear Element)

Neuron sa linearnom aktivacionom funkcijom se naziva linearni neuron. Neka je na raspolaganju obučavajući skup

$$\left\{(x^{(1)}, d^{(1)}), \ldots, (x^{(p)}, d^{(p)})\right\}. \tag{7.1}$$

Cilj obučavanja je izračunavanje težina w_i, koje zadovoljavaju relaciju

$$\sum_{j=1}^{m} w_j x_j^{(k)} = d^{(k)}, \quad k = 1, 2, \ldots, p \tag{7.2}$$

i pri tome se minimizi

$$E(w) = \frac{1}{2}\sum_{k=1}^{p}(d^{(k)} - y^{(k)})^2 = \frac{1}{2}\sum_{k=1}^{p}(d^{(k)} - W^T x^{(k)})^2 = \frac{1}{2}\sum_{k=1}^{p}(d^{(k)} - \sum_{j=1}^{m} w_j x_j^{(k)})^2 .$$

Ekstremizaciju kriterijuma (7.3) možemo obaviti gradijentnom metodom. datom sa

ADALINA (Adaptive Linear Element)

$$\Delta w = \eta \, \nabla_w E(w) , \qquad (7.4)$$

odnosno

$$\Delta w_j = \eta \frac{\partial E}{\partial w_j} = \eta \sum_{k=1}^{p} (d^{(k)} - W^T x^{(k)}) \, x_j^{(k)} , \quad j = 1,2,\ldots,m \qquad (7.5)$$

Ukoliko se ove promene obavijaju individualno za svaki ulazni signal ,
nalazimo da je

$$\Delta w_j = \eta \, (d^{(k)} - W^T x^{(k)}) \, x_j^{(k)} \qquad (7.6)$$

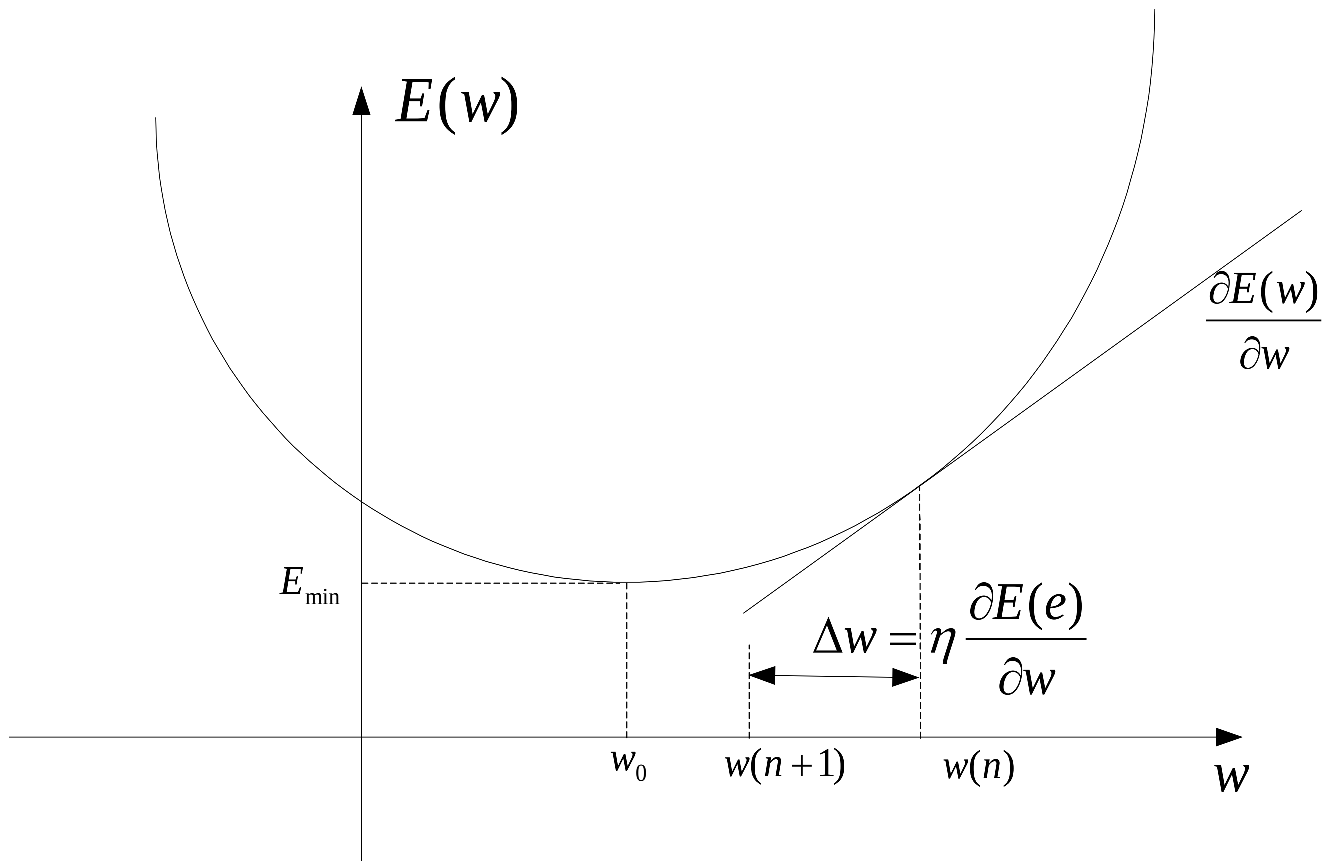
što je poznato Vidrov-Hofovo pravilo obučavanja. Ono se susreće i pod nazivom LMS pravilo (pravilo najmanjih kvadrata, Least Mean Square).

Ako želimo da Vidrov-Hofovo pravilo obučavanja izvedemo iz opšte jednačine obučavanja, neophodno je staviti za signal učenja

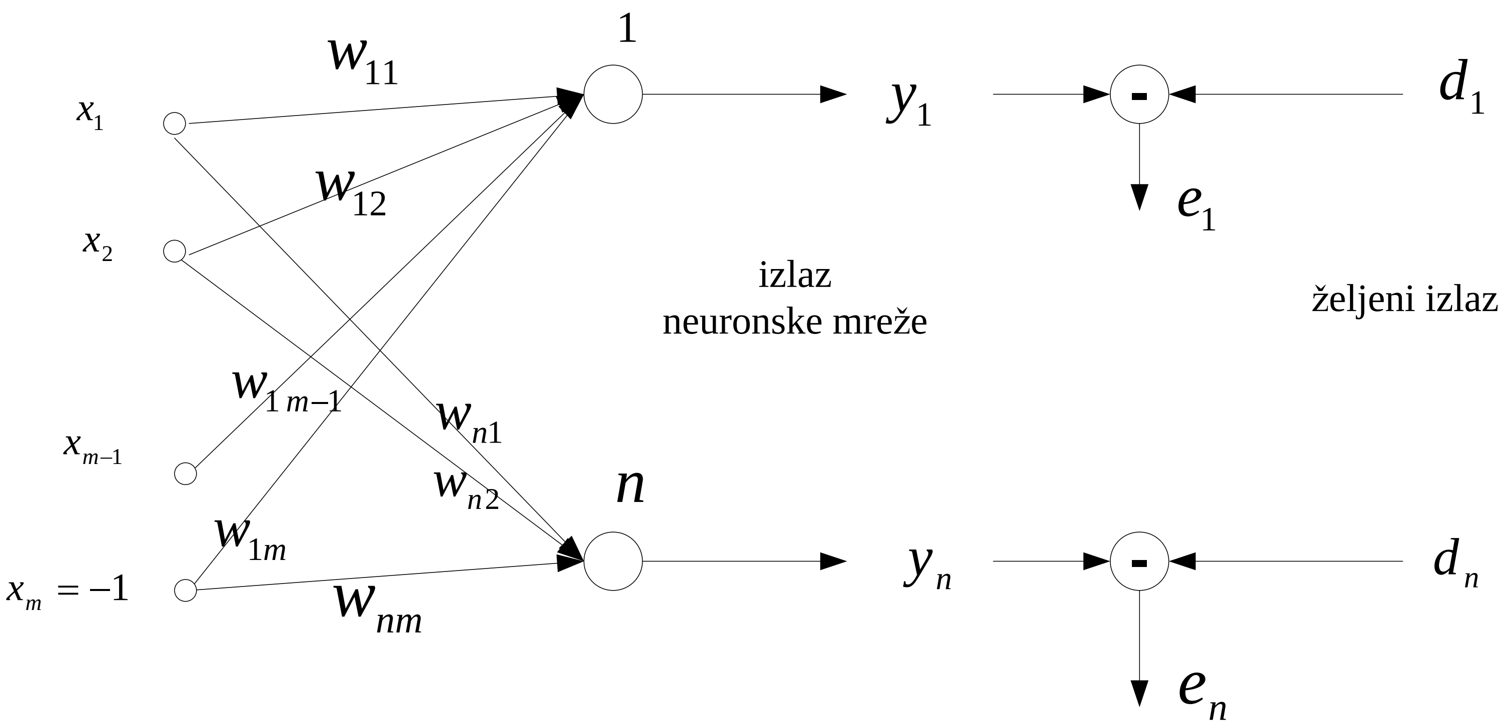
$$r = d - y = d - W^T x . \qquad (7.7)$$

Budući da je E(w) hiperparabolična površ u prostoru sinaptičkih težina w, sa jedinstvenim globalnim ekstremumom (minimumom), postupak konvergira ka njemu bez obzira na početne uslove, pod uslovom da je dovoljno malo.

Ilustracija Vidrov-Hofovog pravila obučavanja za jedan koeficijent sinaptičkih težina w



Jednoslojni perceptron



Jednoslojni perceptron

$\left(w_{1m} = \theta_1,\, w_{2m} = \theta_2,\, ...,\, w_{nm} = \theta_n\right)$

m – broj ulaza

n – brj izlaza

p – dužina obučavajućeg skupa

$$y_i^{(k)} = a\left(W_i^T x^{(k)}\right) = a\left(\sum_{j=1}^{m} w_{ij} x_j^{(k)}\right) = d_i^{(k)}\ , \quad i = 1,2,\ldots,n, \quad k = 1,2,\ldots.p. \quad (8.1)$$

gde je

$$W_i^T = \left[w_{i1}, w_{i2}, \ldots, w_{im}\right]^T \quad (8.2)$$

vektor težina pridružen neuronu i. Ako definišemo kriterijumsku funkciju kao matematičko očekivanje greške na izlazu neuronske mreže, odnosno u slučaju konačnih obučavajućih skupova u obliku ukupne kvadratne greške na obučavajućem skupu, dobijamo

Jednoslojni perceptron

$$E(w) = \frac{1}{2}\sum_{k=1}^{p}\sum_{i=1}^{n}\left(d_i^k - y_i^{(k)}\right)^2 = \frac{1}{2}\sum_{k=1}^{p}\sum_{i=1}^{n}\left[d_i^{(k)} - a(w_i^T x^{(k)})\right]^2 = \frac{1}{2}\sum_{k=1}^{p}\sum_{i=1}^{n}\left[d_i^{(k)} - a\left(\sum_{j=1}^{m} w_{ij} x_j^{(k)}\right)\right]^2.$$

$$\frac{\partial E}{\partial w_{ij}} = -\sum_{k=1}^{p}\left[d_i^{(k)} - a\left(net_i^{(k)}\right)\right]a'\left(net_i^{(k)}\right)x_j^{(k)} \quad , \qquad (8.3)$$

$net_i^{(k)} = W_i^T x^{(k)}$ - ulaz u i-ti neuron kada je k-ti ulazni vektor prisutan.

$$a'\left(net_i^{(k)}\right) = \frac{\partial a\left(net_i^{(k)}\right)}{\partial net_i^{(k)}} \quad . \qquad (8.4)$$

Jednoslojni perceptron

Korekcija w_{ij} nakon prezentacije k-tog obučavajućeg uzorka je

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} = \eta \left[d_i^{(k)} - a(net_i^{(k)}) \right] a' \left(net_i^{(k)} \right) x_j^{(k)} \quad , \qquad (8.5)$$

i naziva se Delta pravilo obučavanja (delta learning rule), koje se iz opšteg pravila obučavanja dobija stavljanjem

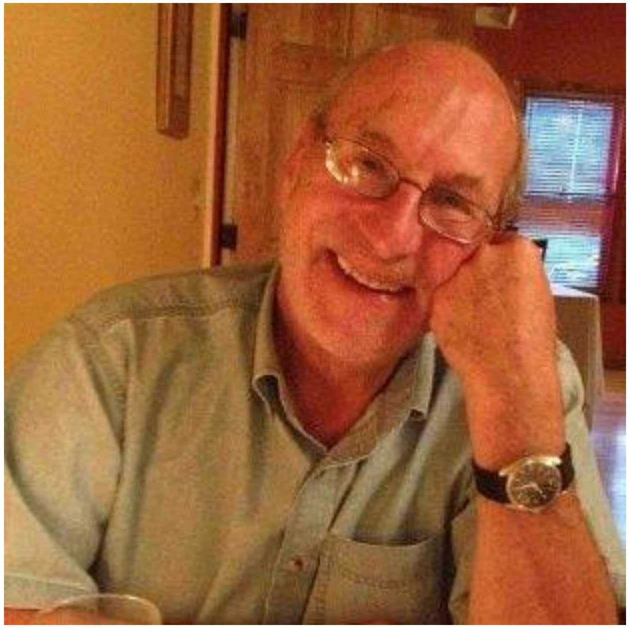
$$r = \left[d_i - a(w_i^T x) \right] a' (w_i^T x) \,. \qquad (8.6)$$

Opisana procedura konvergira ka nekom od lokalnih ekstremuma. Budući da kriterijum obučavanja poseduje više lokalnih ekstremuma, gradijentna procedura (8.5) ne garantuje globalni, već samo neki od lokalnih ekstremuma, zavisno od početnih uslova i parametara obučavanj.

Cibenkova teorema (1989)

- Jedna od najvažnijih karakteristika višeslojnih neuronskih mreža je njihova sposobnost restauracije široke klase ulazno izlaznih preslikavanja, pod relativno opštim uslovima nametnutim aktivacionim funkcijama. Prvi teorijski dokaz ovog važnog stava dao je George Cybenko 1989. god.
- Kurt Hornik je 1991. godine pokazao da dobra aproksimirajuća svojstva višeslojnih perceptrona ne potiču od posebnih svojstava aktivacionih funkcija, već od same arhitekture neuronske mreže.

Cibenkova teorema (1989)



George Cybenko (levo) i Kurt Hornik (desno). Cybenko je prvi pokazao 1989. godine da je jednoslojna neuronska mreža univerzalni aproksimator. Hornik je 1991. godine pokazao da svojstvo univerzalnog aproksimatora ne potiče od posebnih osobina aktivacionih funkcija, već od same arhitekture neuronskih mreža.

Cibenkova teorema (1989)

HSW Teorema

Višeslojna neuronska mreža sa najmanje jednim skrivenim slojem i aktivacionom funkcijom koja poseduje sledeća svojstva

1. $\lim_{\lambda \to \infty} a(\lambda) = 1$

2. $\lim_{\lambda \to -\infty} a(\lambda) = 0 \quad (-1)$

3. $a(\lambda)$ *je neopadajuća funkcija*

aproksimira bilo koju Borel merljivu funkciju na kompaktnim skupovima, sa proizvoljnom tačnošću, pod uslovom da je na raspolaganju dovoljan broj neurona u skrivenom sloju.

Cibenkova teorema (1989)

- Borel merljive funkcije na kompaktnim skupovima obuhvataju sve neprekidne i u delovima neprekidne funkcije (sa konačno ili prebrojivo mnogo diskontinuiteta na skupovima mere nula).

- Odavde sledi da je FFANN univerzalni aproksimator. Stoga neuspeh FFANN da u nekom konkretnom slučaju restauriše preslikavanje implicitno zadato obučavajućim skupom, potiče ili od neadekvatnog izbora arhitekture, parametara obučavanja, obučavajućih skupova i drugih faktora, ali ne i od samog osnovnog restauratorskog principa FFANN.
- Za mnoge praktične probleme, pokazuje se da uprkos HSV teoremi jedan skriveni sloj nije dovoljan, budući da zahteva neprihvatljivo velik broj neurona. raktično bolji rezultati se često dobijaju razmeštanjem manjeg broja neurona u dva ili više skrivenih slojeva.

Algoritam propagacije greške unazad (Back – propagation -BP)

Ovaj algoritam obuhvata dve faze:

1. ulazni vektor $x^{(k)}$ propagira od ulaznog ka izlaznom sloju, produkujući izlaz $y^{(k)}$.
2. signal greške, zatim u drugoj fazi propagira unazad od izlaznog ka ulaznom sloju u cilju korigovanja težina w_{ij}.

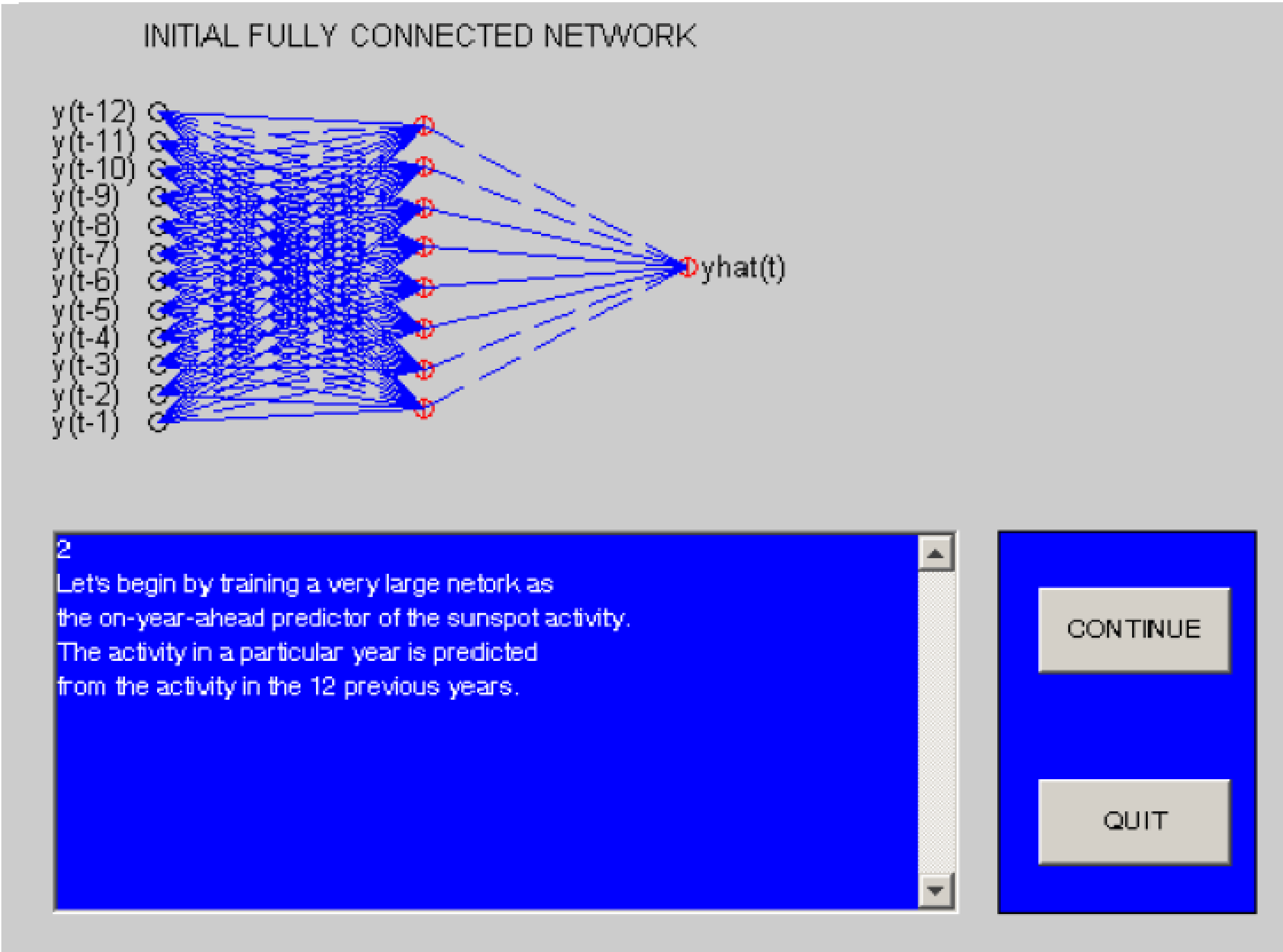
U cilju ilustracije rada algoritma propagacije greške unazad (BP algoriram) razmotrimo višeslojni perceptron tipa m – l – n sa jednim skrivenim slojem.

Problemi konvergencije BP algoritma

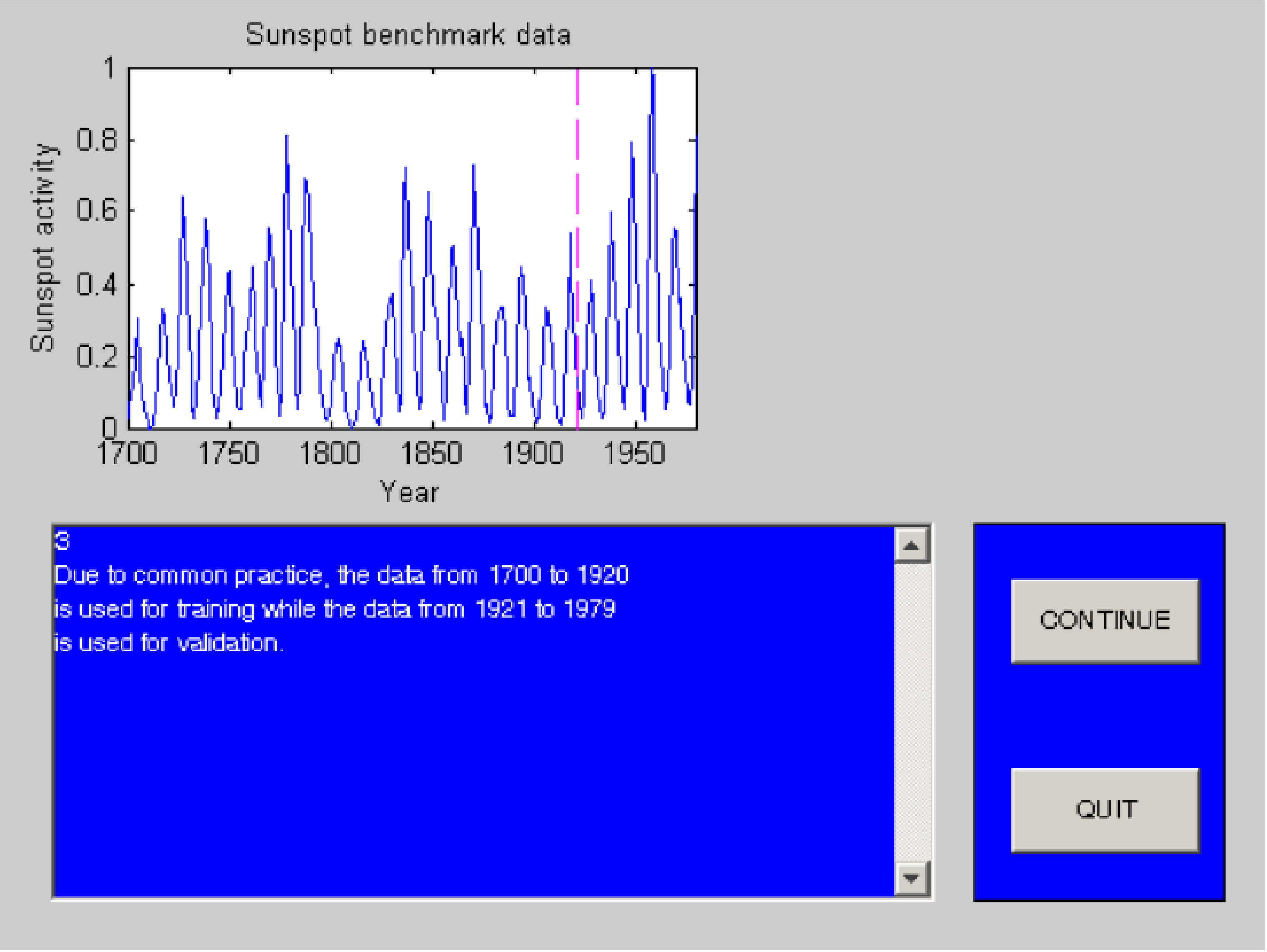
Površina na kojoj se traži ekstremum (error surface – površina greške) nije deterministička. Algoritam ustvari pripada klasi algoritama stohastičke aproksimacije. Za površinu greške se znaju tri bazična svojstva:

- veliki broj lokalnih minimuma, budući da postoji veliki broj kombinatornih permutacija težina koje daju isti izlaz mreže.
- postojanje lokalnih minimuma iznad nivoa globalnog minimuma
- postojanje višestrukih platoa sa malim nagibima. Ovo je direktna posledica zasićenja aktivacionih funkcija u domenu velikih signala, kada su izlazi neosetljivi na male promene težina. Postojanje ovakvih delova površine greške prouzrokuje sporu konvergenciju algoritma propagacije greške unazad.

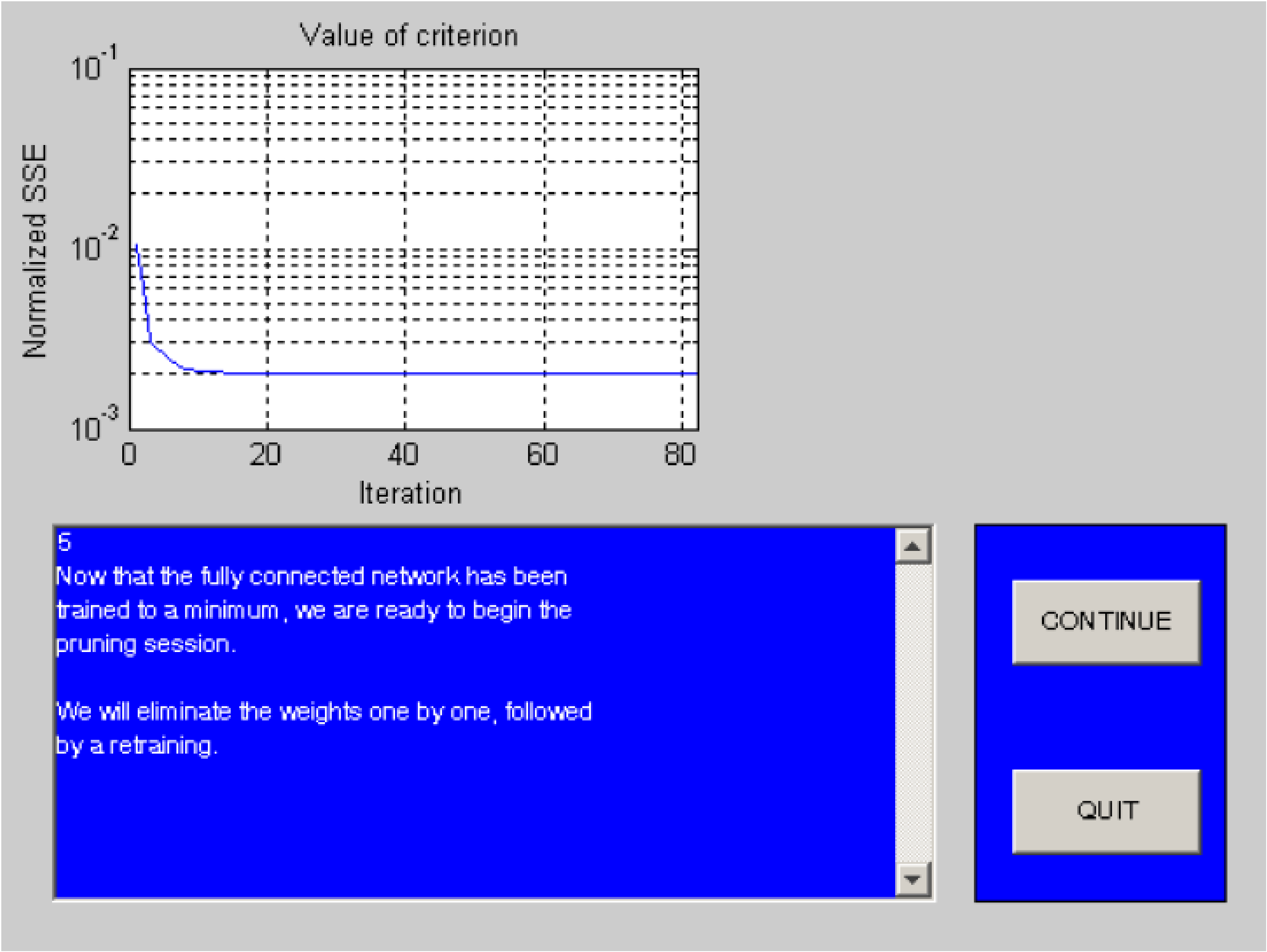
Primer



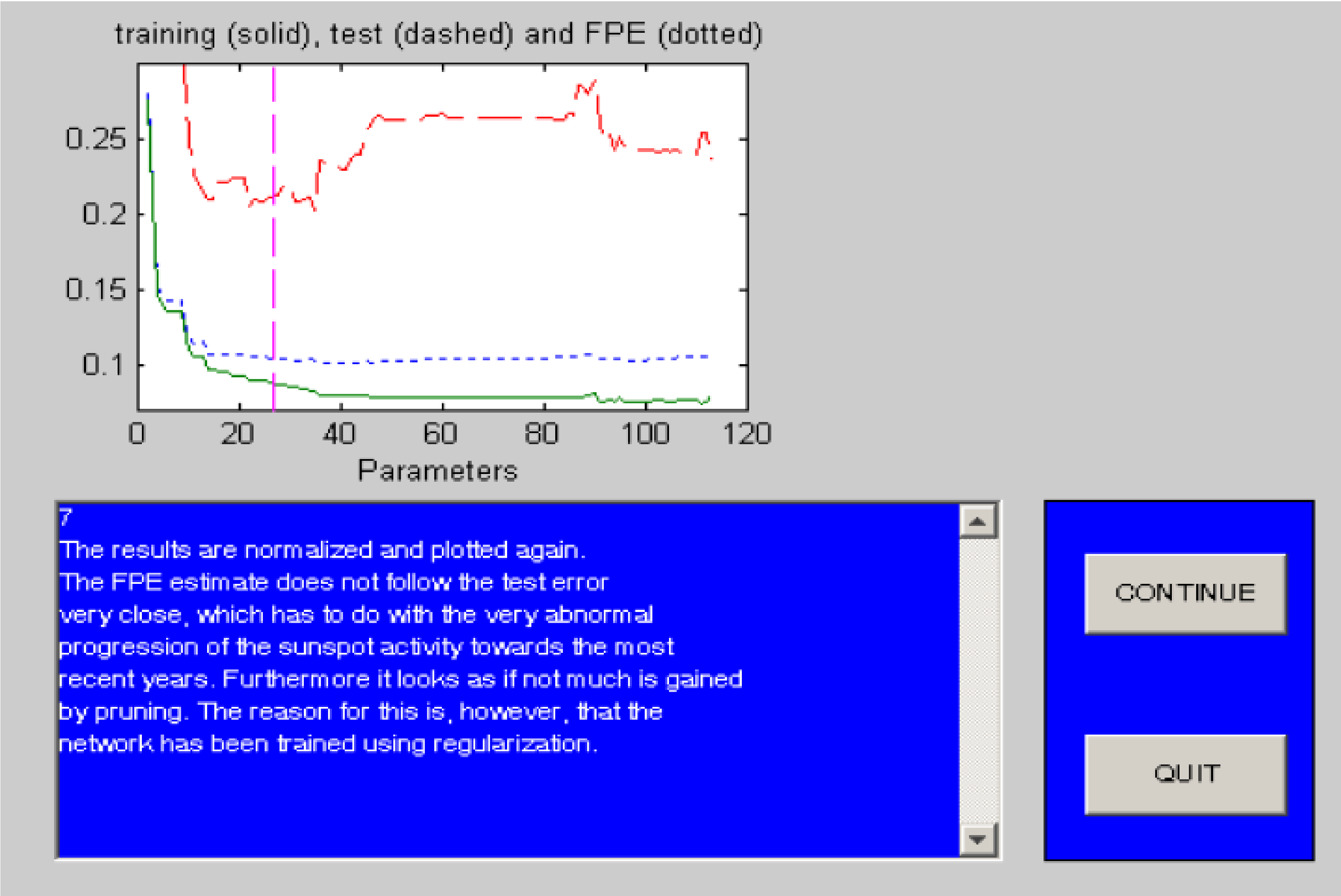
Početna arhitektura neuronskog prediktora 12. reda, budući da koristi 12 prošlih godina za predvidjanje sunčevih pega u tekućoj godini. Primenjena neuronska mreža ima 12 ulaza, 8 neurona u skrivenom sloju i jedan izlazni neuron. Po svojoj prirodi to je jednoslojni perceptron sa prostiranjem signala unapred.



Prikaz analizirane vremenske serije broja sunčanih pega u periodu (1700-1979). Posebno je označen interval treninga (1700-1920) i validacije (1921-1979).



Prikaz promene greške treninga u toku obučavanja početne neuronske mreže sa maksimalnim brojem sinaptičkih veza.



Prikaz tri vrste greške u toku procesa kresanja pune arhitekture: greška na treningu (puna zelena linija), greška na test skupu (isprekidana crvena linija), procena finalne greške predikcije (Final Prediction Error –FPE) u funkciji broja parametara neuronske mreže. Procena FPE daje u stvari procenu greške generalizacije.