Neizrazito, evolucijsko i neuro računarstvo: 6. domaća zadaća - izvještaj

Krešimir Topolovec 0036485747

Siječanj 2020.

1 ANFIS

Koristimo Neuro-Fuzzy sustav ANFIS (engl.Adaptive Neuro-Fuzzy Inference System) koji pripada pod hibridni pristup kombiniranja neuronskih mreža i sustava neizrazitog zaključivanja. Kod ovog sustava neuronska mreža je istovremeno i sustav neizrazitog zaključivanja. U ovoj vježbi koristimo primjer takvog sustava koji ima 2 ulaza i 1 izlaz odnosno obavlja preslikavanje $R^2 \to R$ i koristi zaključivanje tipa 3 (metoda Takagi-Sugeno-Kang).
Prikaz takvog sustava je na slijedećoj slici:

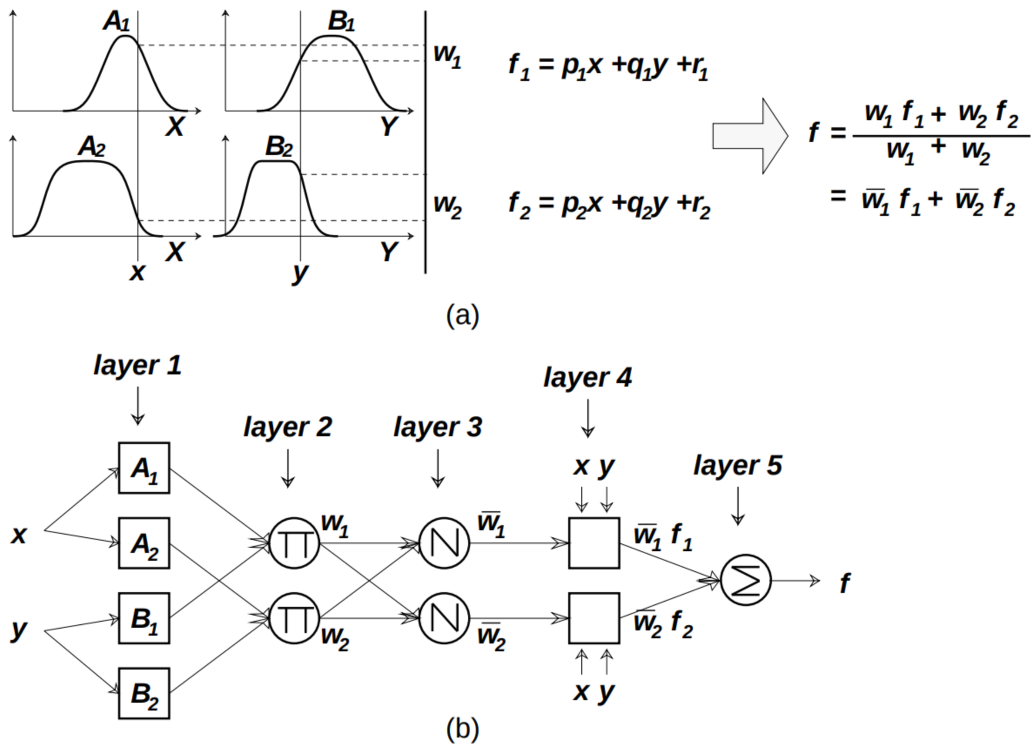

Figure 1: ANFIS sa 2 ulaza i 1 izlazom koji koristi TSK zaključivanje

ANFIS obavlja preslikavanje koristeći pravila oblika:
R_i : AKO x je A_i I y je B_i ONDA $z_i = p_i x + q_i y + r_i$, $i = 1..m$

Neuroni u prvom sloju na slici 1 modeliraju neizrazite skupove. Njihova prijenosna funkcija (a time i funkcija pripadnosti koju koristi sustav ANFIS) je:

$$A_i(x) \equiv \mu_{A_i}(x) = \frac{1}{1 + e^{b_i(x-a_i)}}$$

$$B_i(y) \equiv \mu_{B_i}(y) = \frac{1}{1 + e^{d_i(y-c_i)}}$$

Neuroni drugo sloja računaju konjukciju pravila (ukupnu "jakost" i-tog pravila) pa koristimo t-normu. U zadatku je kao t-norma definiran produkt:

$$\mu_{A_i \cap B_i}(x) = \mu_{A_i}(x) \cdot \mu_{A_i}(x)$$

Izraz μ_{A_i} označimo s α_i, te μ_{B_i} s β_i, jakost pravila i tada je:

$$w_i = \alpha_i \cdot \beta_i \tag{1}$$

Jakost svakog pravila se skalira(dijeli) sumom jakosti ostalih pravila, to obavlja sloj 3. Konsekvent svakog pravila koji je definiran kao funkcija z_i(na slici f_i):

$$z_i = p_i x + q_i y + r_i, i = 1..m \tag{2}$$

množi se sa skaliranom težinom, ovo računa 4. sloj mreže. Izlaz ANFISA sastoji se od sume izlaza neurona 4. sloja i definiran je sa funkcijom:

$$o_k = \frac{\sum_{i=1}^{m} w_i z_i}{\sum_{i=1}^{m} w_i},$$

2 Gradijenti spust - postupak učenja

Zadatak je naučiti hibridnu mrežu ANFIS kako bi računala funkciju:

$$f(x, y) = ((x-1)^2 + (y+2)^2 - 5xy + 3) \cdot cos^2(\frac{x}{5}), \; x \in [-4, 4], y \in [-4, 4].$$

Definiramo kvadratnu pogrešku mreže za k-ti ulazni primjerak gdje je primjerak sastavljen od trojke (x_{1k}, x_{2k}, y_k):

$$E_k = \frac{1}{2}(y_k - o_k)^2 \tag{3}$$

Učenje tj. optimiranje mreže provodimo gradijentnim spustom. Opća formula gradijentnog spusta gdje optimiramo paramentar ψ o kojem greška ovisi je:

$$\psi(t+1) = \psi(t) - \eta \frac{\partial E_k}{\partial \psi} \tag{4}$$

U našoj mreži s 2 ulaza i 1 izlazom parametri o kojima ovisi greška E_k su: $a_i, b_i, c_i, d_i, p_i, q_i$ i r_i Moramo izračunati 7 parcijalnih derivacija funkcije E_k u ovisnosti ovim paremetrima. To su parcijalne derivacije:

$$\frac{\partial E_k}{\partial a_i}, \frac{\partial E_k}{\partial b_i}, \frac{\partial E_k}{\partial c_i}, \frac{\partial E_k}{\partial d_i}, \frac{\partial E_k}{\partial p_i}, \frac{\partial E_k}{\partial q_i}, \frac{\partial E_k}{\partial r_i}$$

Za izvod svake parcijalne derivacije koristimo pravilo ulančavanja. Npr. za izračun parcijalne derivacije za parametar a_i koristimo lanac:

$$E_k \to o_k \to w_i \to \alpha_i \to a_i$$

odnosno:

$$\frac{\partial E_k}{\partial a_i} = \frac{\partial E_k}{\partial o_k} \cdot \frac{\partial o_k}{\partial w_i} \cdot \frac{\partial w_i}{\partial \alpha_i} \cdot \frac{\partial \alpha_i}{\partial a_i} \tag{5}$$

Analogno i za sve ostale paremetre mreže:

$$\frac{\partial E_k}{\partial b_i} = \frac{\partial E_k}{\partial o_k} \cdot \frac{\partial o_k}{\partial w_i} \cdot \frac{\partial w_i}{\partial \alpha_i} \cdot \frac{\partial \alpha_i}{\partial b_i} \tag{6}$$

$$\frac{\partial E_k}{\partial c_i} = \frac{\partial E_k}{\partial o_k} \cdot \frac{\partial o_k}{\partial w_i} \cdot \frac{\partial w_i}{\partial \beta_i} \cdot \frac{\partial \beta_i}{\partial c_i} \tag{7}$$

$$\frac{\partial E_k}{\partial d_i} = \frac{\partial E_k}{\partial o_k} \cdot \frac{\partial o_k}{\partial w_i} \cdot \frac{\partial w_i}{\partial \beta_i} \cdot \frac{\partial \beta_i}{\partial d_i} \tag{8}$$

$$\frac{\partial E_k}{\partial p_i} = \frac{\partial E_k}{\partial o_k} \cdot \frac{\partial o_k}{\partial z_i} \cdot \frac{\partial z_i}{\partial p_i} \tag{9}$$

$$\frac{\partial E_k}{\partial q_i} = \frac{\partial E_k}{\partial o_k} \cdot \frac{\partial o_k}{\partial z_i} \cdot \frac{\partial z_i}{\partial q_i} \tag{10}$$

$$\frac{\partial E_k}{\partial r_i} = \frac{\partial E_k}{\partial o_k} \cdot \frac{\partial o_k}{\partial z_i} \cdot \frac{\partial z_i}{\partial r_i} \tag{11}$$

Možemo uočiti da se mnogi članovi često ponavljaju što olakšava implementaciju jer ih ne moramo višestruko računati nego jednom za svaki ulazni uzorak k Članovi parcijalnih derivacija koji se ponavljaju su: $\frac{\partial E_k}{\partial o_k}$, $\frac{\partial o_k}{\partial w_i}$, $\frac{\partial o_k}{\partial z_i}$, $\frac{\partial w_i}{\partial \alpha_i}$ i $\frac{\partial w_i}{\partial \beta_i}$

Izračun parcijalne derivacije: $\frac{\partial \mathbf{E_k}}{\partial \mathbf{o_k}}$

$$\frac{\partial E_k}{\partial o_k} = \frac{\partial}{\partial o_k}\left(\frac{1}{2}(y_k - o_k)^2\right) = -(y_k - o_k) \tag{12}$$

Izračun parcijalne derivacije: $\frac{\partial \mathbf{o_k}}{\partial \mathbf{w_i}}$

$$\begin{aligned} \frac{\partial o_k}{\partial w_i} &= \frac{\partial}{\partial w_i}\left(\frac{\sum_{j=1}^m w_j z_j}{\sum_{j=1}^m w_j}\right) \\ &= \frac{\partial}{\partial w_i}\left(\frac{w_1 z_1 + w_2 z_2 + ... + w_i z_i + ... + w_m z_m}{\sum_{j=1}^m w_j}\right) \end{aligned}$$

Koristimo pravilo za derivaciju razlomka...

$$= \frac{\partial}{\partial w_i}\left(\frac{w_1 z_1}{\sum_{j=1}^m w_j}\right) + ... + \frac{\partial}{\partial w_i}\left(\frac{w_i z_i}{\sum_{j=1}^m w_j}\right) + ... + \frac{\partial}{\partial w_i}\left(\frac{w_m z_m}{\sum_{j=1}^m w_j}\right) \tag{13}$$

$$= \frac{-w_1 z_1}{(\sum_{j=1}^m w_j)^2} + ... + \frac{z_i \sum_{j=1}^m w_j + w_i z_i}{(\sum_{j=1}^m w_j)^2} + ... + \frac{-w_m z_m}{(\sum_{j=1}^m w_j)^2} \tag{14}$$

$$= \frac{-w_1 z_1}{(\sum_{j=1}^m w_j)^2} + ... + \frac{-w_i z_i}{(\sum_{j=1}^m w_j)^2} + ... + \frac{-w_m z_m}{(\sum_{j=1}^m w_j)^2} + \frac{z_i \sum_{j=1}^m w_j}{(\sum_{j=1}^m w_j)^2} \tag{15}$$

$$\frac{\partial o_k}{\partial w_i} = \frac{\sum_{j=1}^m w_j \cdot (z_i - z_j)}{(\sum_{j=1}^m w_j)^2} \tag{16}$$

Izračun parcijalne derivacije $\frac{\partial \mathbf{o_k}}{\partial \mathbf{z_i}}$

$$\frac{\partial o_k}{\partial z_i} = \frac{\partial}{\partial z_i}\left(\frac{\sum_{i=1}^m w_i z_i}{\sum_{i=1}^m w_i}\right)$$

$$\frac{\partial o_k}{\partial z_i} = \frac{w_i}{\sum_{j=1}^m w_j} \tag{17}$$

Izračun parcijalne derivacije $\frac{\partial \mathbf{w_i}}{\partial \beta_\mathbf{i}}$ i $\frac{\partial \mathbf{w_i}}{\partial \alpha_\mathbf{i}}$

$$\frac{\partial w_i}{\partial \beta_i} = \alpha_i$$
$$\frac{\partial w_i}{\partial \alpha_i} = \beta_i$$

Izračun parcijalne derivacije za parametre a_i, b_i, c_i te d_i

$$\frac{\partial E_k}{\partial a_i} = \frac{\partial E_k}{\partial o_k} \cdot \frac{\partial o_k}{\partial w_i} \cdot \frac{\partial w_i}{\partial \alpha_i} \cdot \frac{\partial \alpha_i}{\partial a_i}$$

Prva tri člana u ovo izrazu smo već dobili iznad, potrebna nam je još parcijalna derivacija α_i po a_i.

$$\frac{\partial \alpha_i}{\partial a_i} = \frac{\partial}{\partial a_i}\left(\frac{1}{1+e^{b_i(x-a_i)}}\right) = \frac{\partial \alpha_i}{\partial a_i} = \alpha_i \cdot (1-\alpha_i) \cdot b_i$$

Izraz za ažuriranje parametra a_i konačno je:

$$a_i(t+1) = a_i(t) + \eta \cdot (y_k - o_k) \cdot \frac{\sum_{j=1}^m w_j \cdot (z_i - z_j)}{(\sum_{j=1}^m w_j)^2} \cdot \beta_i \cdot \alpha_i \cdot (1-\alpha_i) \cdot b_i \tag{18}$$

$$\frac{\partial E_k}{\partial b_i} = \frac{\partial E_k}{\partial o_k} \cdot \frac{\partial o_k}{\partial w_i} \cdot \frac{\partial w_i}{\partial \alpha_i} \cdot \frac{\partial \alpha_i}{\partial b_i}$$

$$\frac{\partial \alpha_i}{\partial b_i} = \frac{\partial}{\partial b_i}\left(\frac{1}{1+e^{b_i(x-a_i)}}\right) = \alpha_i \cdot (1-\alpha_i) \cdot (x-a_i)$$

Izraz za ažuriranje parametra b_i konačno je:

$$b_i(t+1) = b_i(t) + \eta \cdot (y_k - o_k) \cdot \frac{\sum_{j=1}^m w_j \cdot (z_i - z_j)}{(\sum_{j=1}^m w_j)^2} \cdot \beta_i \cdot \alpha_i \cdot (1-\alpha_i) \cdot (x-a_i) \tag{19}$$

Analogni postupak za parametre c_i i d_i

$$c_i(t+1) = c_i(t) + \eta \cdot (y_k - o_k) \cdot \frac{\sum_{j=1}^m w_j \cdot (z_i - z_j)}{(\sum_{j=1}^m w_j)^2} \cdot \alpha_i \cdot \beta_i \cdot (1-\beta_i) \cdot di \tag{20}$$

$$d_i(t+1) = d_i(t) - \eta \cdot (y_k - o_k) \cdot \frac{\sum_{j=1}^m w_j \cdot (z_i - z_j)}{(\sum_{j=1}^m w_j)^2} \cdot \alpha_i \cdot \beta_i \cdot (1-\beta_i) \cdot (x-c_i) \tag{21}$$

Izračun parcijalne derivacije za parametre p_i, q_i te r_i

Preostalo je još izračunati parcijalne derivacije po parametrima p_i, q_i te r_i Za to su nam potrebne parcijalne derivacije $\frac{\partial z_i}{\partial p_i}$, $\frac{\partial z_i}{\partial q_i}$ i $\frac{\partial z_i}{\partial r_i}$

$$\frac{\partial z_i}{\partial p_i} = \frac{\partial}{\partial p_i}(p_i x + q_i y + r_i) = x$$
$$\frac{\partial z_i}{\partial q_i} = \frac{\partial}{\partial q_i}(p_i x + q_i y + r_i) = y$$
$$\frac{\partial z_i}{\partial r_i} = \frac{\partial}{\partial r_i}(p_i x + q_i y + r_i) = 1$$

Konačni izrazi za ažuriranje parametara p_i, q_i te r_i

$$p_i(t+1) = p_i(t) + \eta \cdot (y_k - o_k) \cdot \frac{w_i}{\sum_{j=1}^{m} w_j} \cdot x \tag{22}$$

$$q_i(t+1) = q_i(t) + \eta \cdot (y_k - o_k) \cdot \frac{w_i}{\sum_{j=1}^{m} w_j} \cdot y \tag{23}$$

$$r_i(t+1) = r_i(t) + \eta \cdot (y_k - o_k) \cdot \frac{w_i}{\sum_{j=1}^{m} w_j} \tag{24}$$

Učenje na cijelom skupu za učenje

Definiramo srednju kvadratnu pogrešku mreže za svih k ulaznih primjeraka gdje je primjerak sastavljen od para (x_{1k}, x_{2k}, y_k) a izlaz mreže o_k, a N broj primjeraka za učenje:

$$E = \frac{1}{2N} \sum_{k=1}^{N} (y_k - o_k)^2, \tag{25}$$

Moramo odrediti kako se mijenja pogreška E u ovisnosti o izlazu mreže o

$$\begin{aligned}
\frac{\partial E}{\partial o} &= \frac{\partial}{\partial o_k} \left(\frac{1}{2N} \sum_{k=1}^{N} (y_k - o_k)^2 \right) \\
&= \frac{1}{2N} \sum_{k=1}^{N} \left(\frac{\partial}{\partial o_k} (y_k - o_k)^2 \right) \\
&= \frac{1}{2N} \sum_{k=1}^{N} \left(-2(y_k - o_k) \right) \\
&= -\frac{1}{N} \sum_{k=1}^{N} (y_k - o_k)
\end{aligned}$$

Ostali članovi parcijalnih derivacija u lancu se ni za jedan parametar ne mjenjaju. Moramo u prethodno izvedenim izrazima (18) - (24) promjeniti samo član $\frac{\partial E_k}{\partial o_k}$ sa novoizvedenim $\frac{\partial E}{\partial o}$

$$a_i(t+1) = a_i(t) + \eta \cdot \frac{1}{N} \sum_{k=1}^{N} (y_k - o_k) \cdot \frac{\sum_{j=1}^{m} w_j \cdot (z_i - z_j)}{(\sum_{j=1}^{m} w_j)^2} \cdot \beta_i \cdot \alpha_i \cdot (1 - \alpha_i) \cdot b_i \tag{26}$$

$$b_i(t+1) = b_i(t) - \eta \cdot \frac{1}{N} \sum_{k=1}^{N} (y_k - o_k) \cdot \frac{\sum_{j=1}^{m} w_j \cdot (z_i - z_j)}{(\sum_{j=1}^{m} w_j)^2} \cdot \beta_i \cdot \alpha_i \cdot (1 - \alpha_i) \cdot (x - a_i) \tag{27}$$

$$c_i(t+1) = c_i(t) + \eta \cdot \frac{1}{N} \sum_{k=1}^{N} (y_k - o_k) \cdot \frac{\sum_{j=1}^{m} w_j \cdot (z_i - z_j)}{(\sum_{j=1}^{m} w_j)^2} \cdot \alpha_i \cdot \beta_i \cdot (1 - \beta_i) \cdot di \tag{28}$$

$$d_i(t+1) = d_i(t) - \eta \cdot \frac{1}{N} \sum_{k=1}^{N} (y_k - o_k) \cdot \frac{\sum_{j=1}^{m} w_j \cdot (z_i - z_j)}{(\sum_{j=1}^{m} w_j)^2} \cdot \alpha_i \cdot \beta_i \cdot (1 - \beta_i) \cdot (x - c_i) \tag{29}$$

$$p_i(t+1) = p_i(t) + \eta \cdot \frac{1}{N} \sum_{k=1}^{N} (y_k - o_k) \cdot \frac{w_i}{\sum_{j=1}^{m} w_j} \cdot x \tag{30}$$

$$q_i(t+1) = q_i(t) + \eta \cdot \frac{1}{N} \sum_{k=1}^{N} (y_k - o_k) \cdot \frac{w_i}{\sum_{j=1}^{m} w_j} \cdot y \tag{31}$$

$$r_i(t+1) = r_i(t) + \eta \cdot \frac{1}{N} \sum_{k=1}^{N} (y_k - o_k) \cdot \frac{w_i}{\sum_{j=1}^{m} w_j} \tag{32}$$

3 Učenje parametara

Funkcija 2 koju učimo je definiran prikazana je na slici:

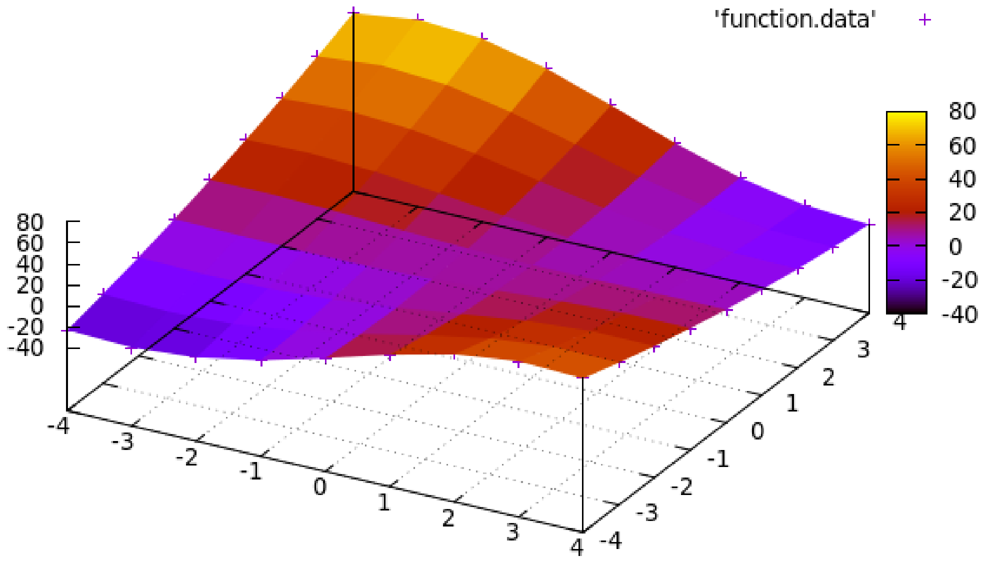

Figure 2: Prikaz funkcije u domeni [-4,4]x[-4,4]

Učenje s 1 pravilom

GD (engl. Gradint Descent) - mormalni gradijentni algoritam
SGD (engl. Stohastic Gradint Descent) - stohastički gradijentni algoritam

GD

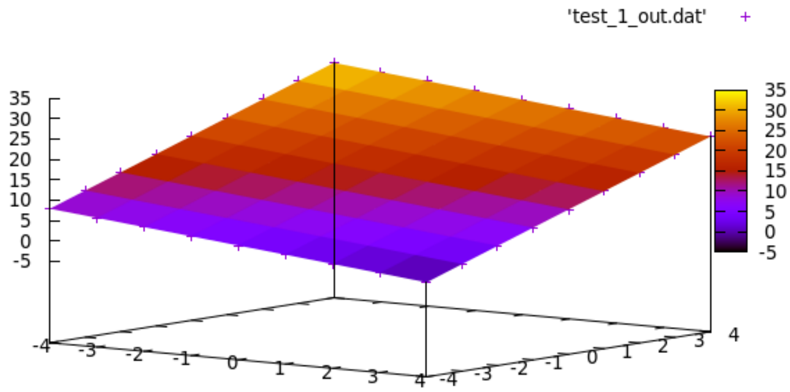

Figure 3: GD Prikaz naučene funkcije za sustav s 1 pravilom

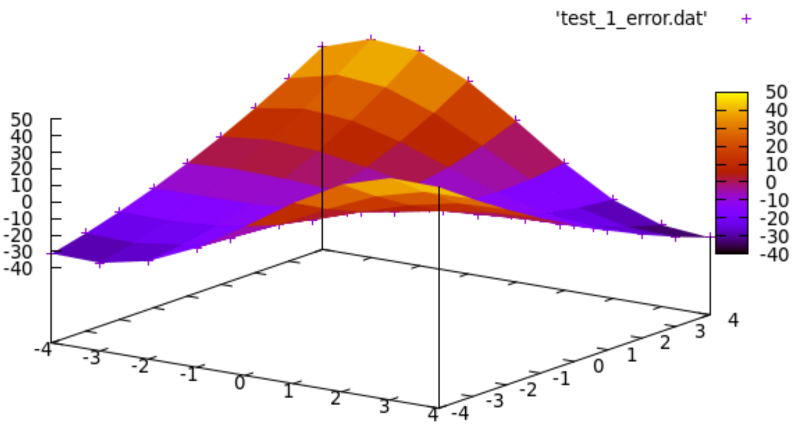

Figure 4: GD Prikaz funkcije pogreške za sustav s 1 pravilom

SGD

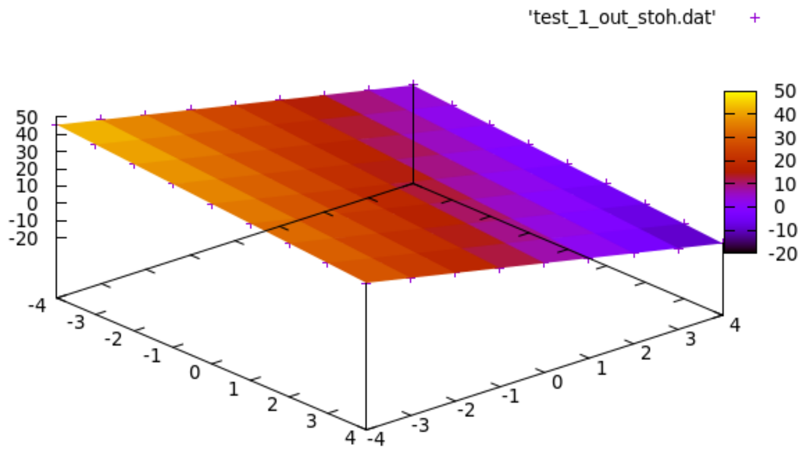

Figure 5: SGD Prikaz naučene funkcije za sustav s 1 pravilom

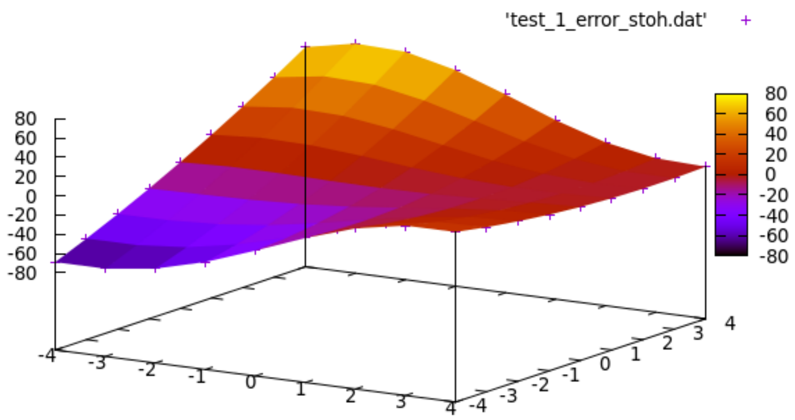

Figure 6: SGD Prikaz funkcije pogreške za sustav s 1 pravilom

Učenje s 2 pravila

GD

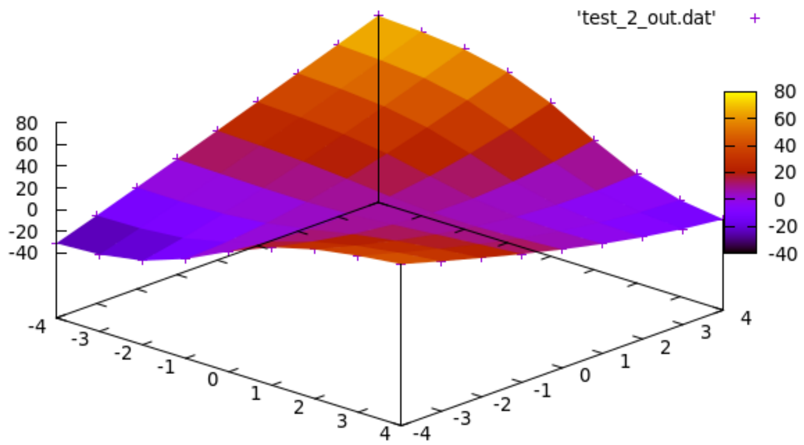

Figure 7: GD Prikaz naučene funkcije za sustav s 2 pravila

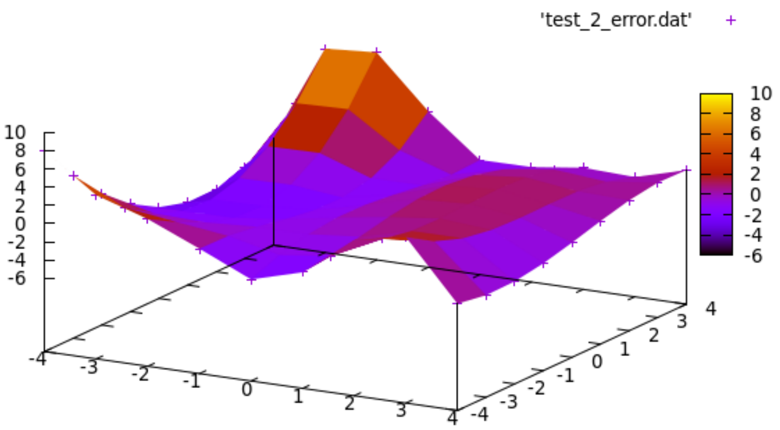

Figure 8: GD Prikaz funkcije pogreške za sustav s 2 pravila

SGD

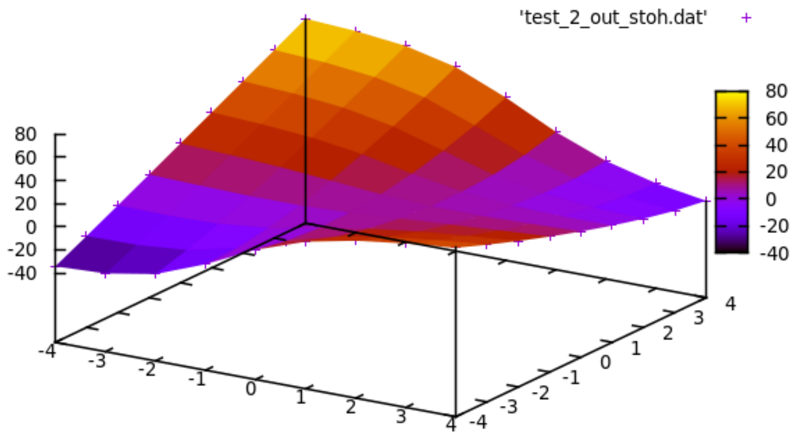

Figure 9: SGD Prikaz naučene funkcije za sustav s 2 pravila

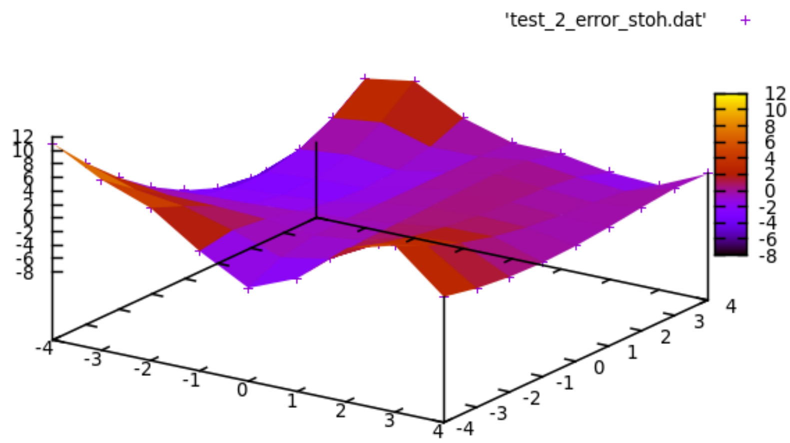

Figure 10: SGD Prikaz funkcije pogreške za sustav s 2 pravila

Učenje s 5 pravila

GD

η_1 : 0.0007, η_2 : 0.01, $error$: 0.08709604459612692

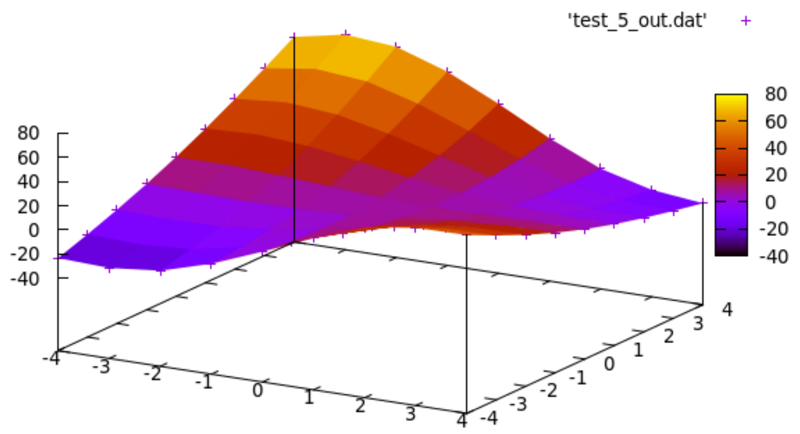

Figure 11: GD Prikaz naučene funkcije za sustav s 5 pravila

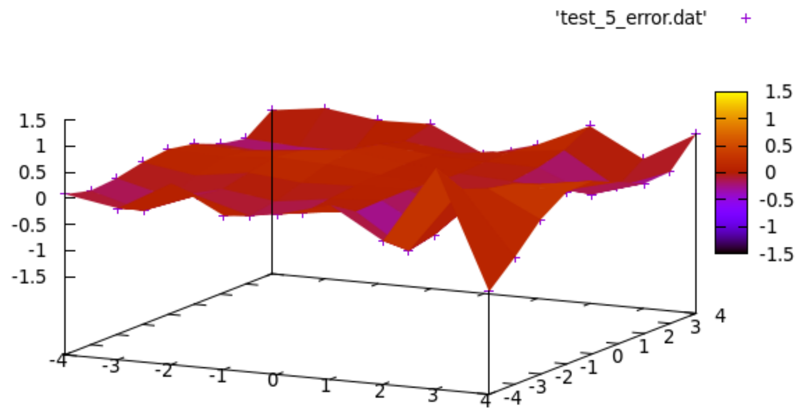

Figure 12: GD Prikaz funkcije pogreške za sustav s 5 pravila

SGD

η_1 : 0.0007, η_2 : 0.01, $error$: 0.030076224714011102

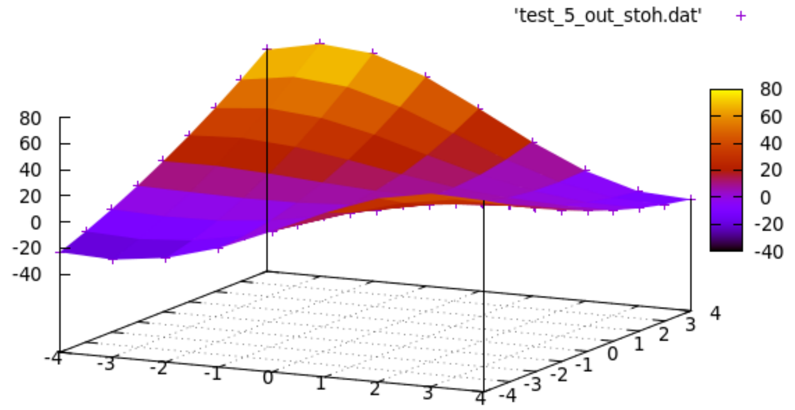

Figure 13: SGD Prikaz naučene funkcije za sustav s 5 pravila

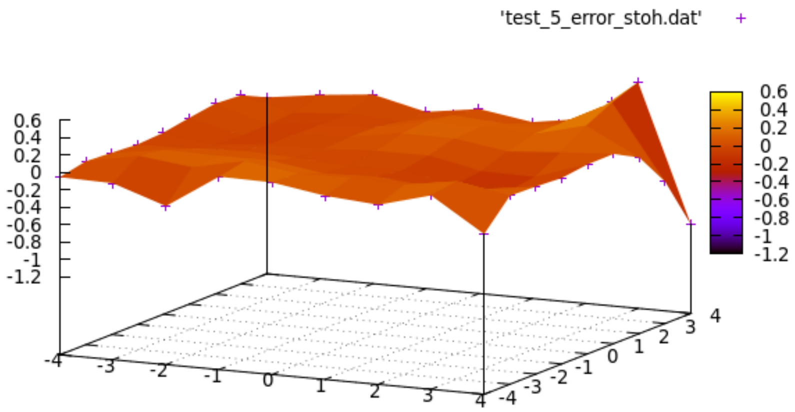

Figure 14: SGD Prikaz funkcije pogreške za sustav s 5 pravila

Grafovi funkcija pripadnosti naučenog sustava

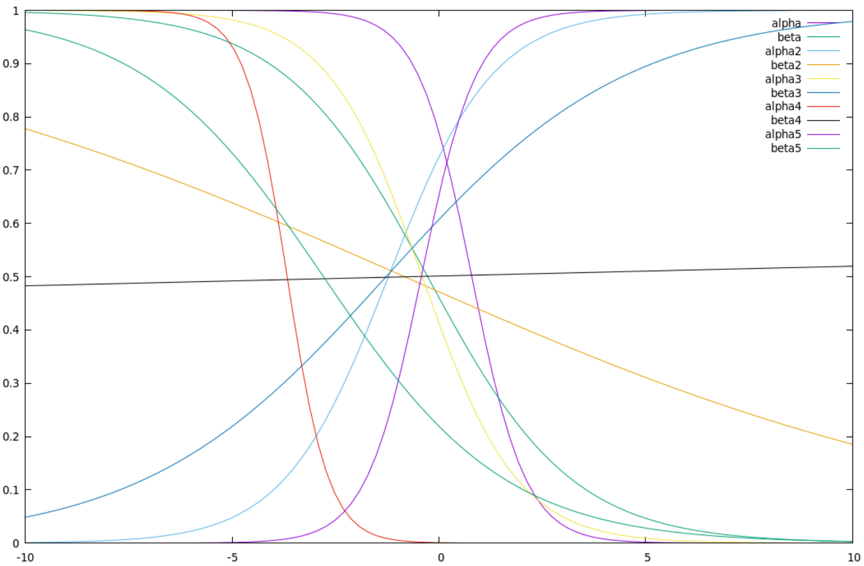

Figure 15: GD Prikaz funkcija pripadnosti

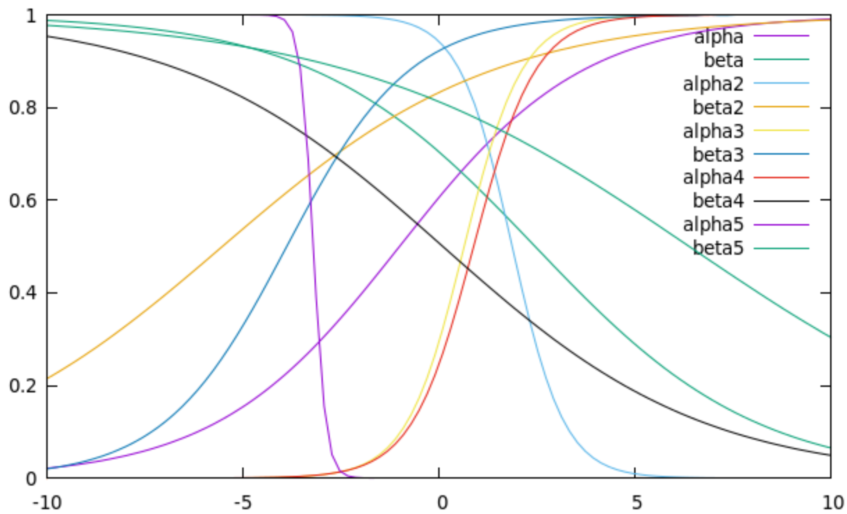

Figure 16: SGD Prikaz funkcija pripadnosti

Kretanje pogreške u ovisnosti o broju epoha

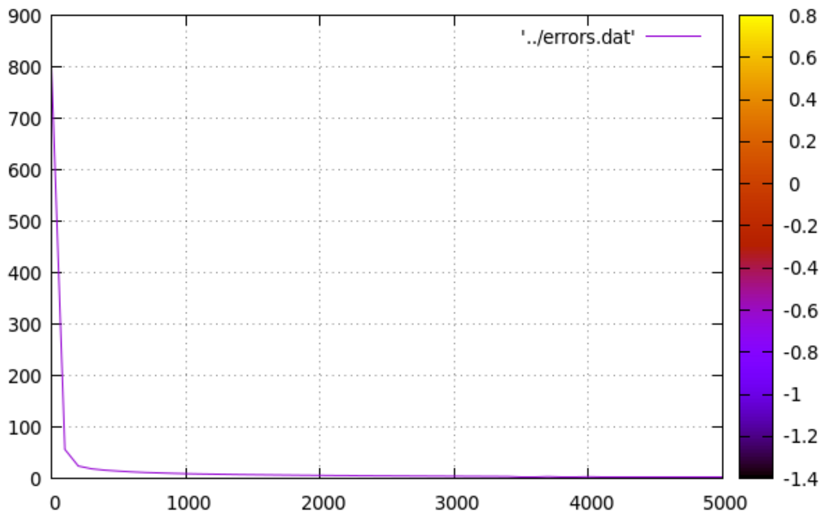

Figure 17: GD

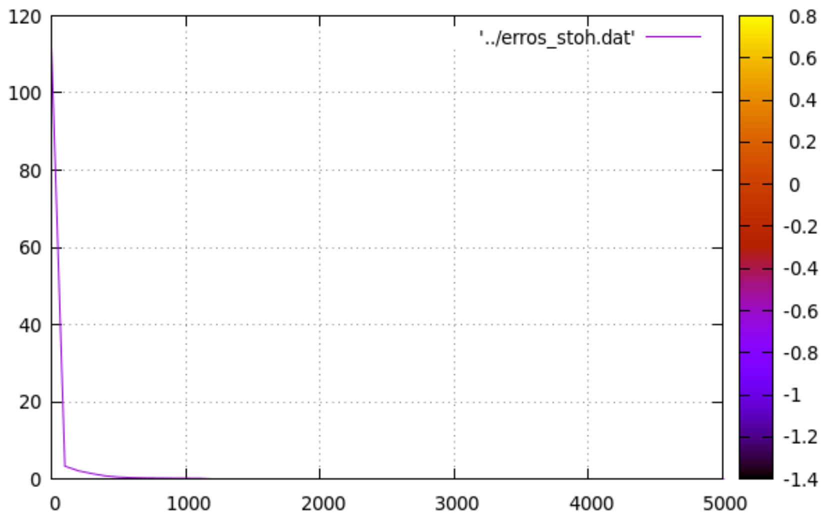

Figure 18: SGD

Kretanje pogreške u ovisnosti o broju epoha za rubne vrijednosti stope učenja

Promjena parametara a_i, b_i, c_i te d_i znatno utječe na strmost sigmoidalne funkcije pa zbog toga te parametre učimo sa stopom učenja manjom (η_1) nego za parametre izlaza (η_2).

Za prevelike stope učenja sustav divergira.
Primjer: GD: $\eta_1 : 0.5$, $\eta_2 : 0.7$, $error : error : 6.158311311006211e + 51$

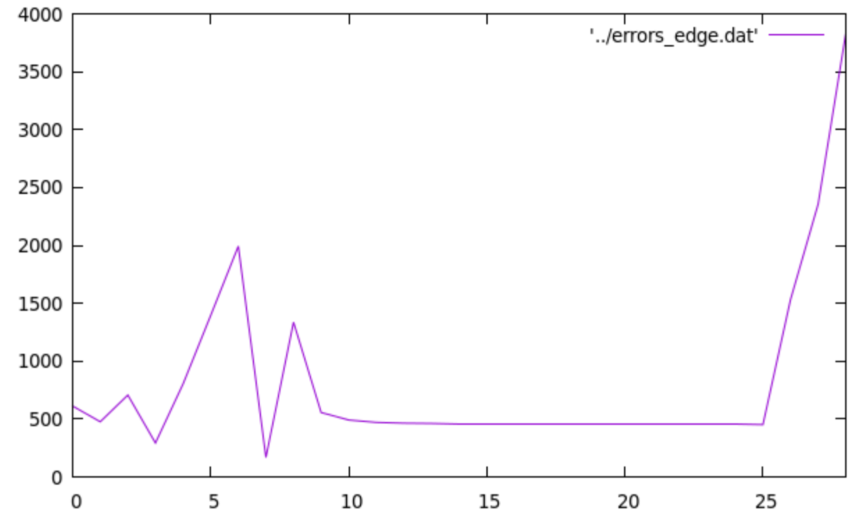

Figure 19: GD Prikaz greške za prevelike stope učenja, domena $[0, 30]$

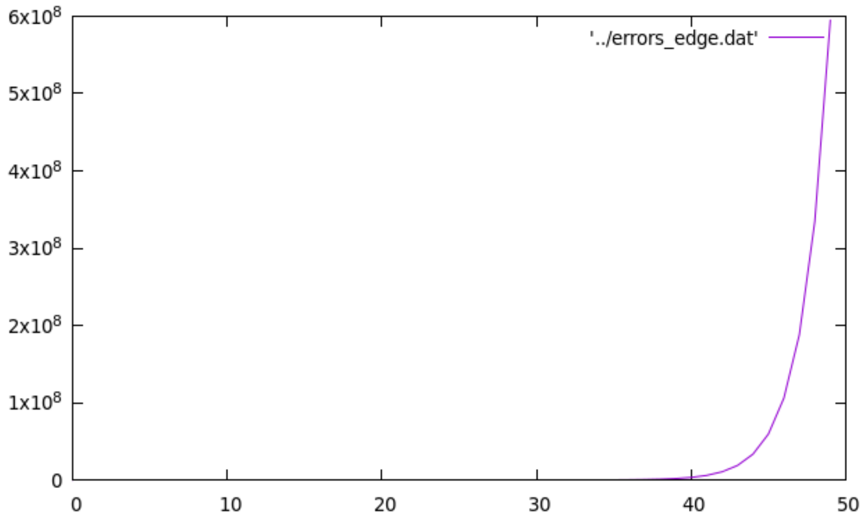

Figure 20: GD Prikaz greške za prevelike stope učenja, domena $[0, 50]$

Primjer: SGD:η_1 : 0.06, η_2 : 0.17, $error$: $error$: $6.158311311006211e + 51$

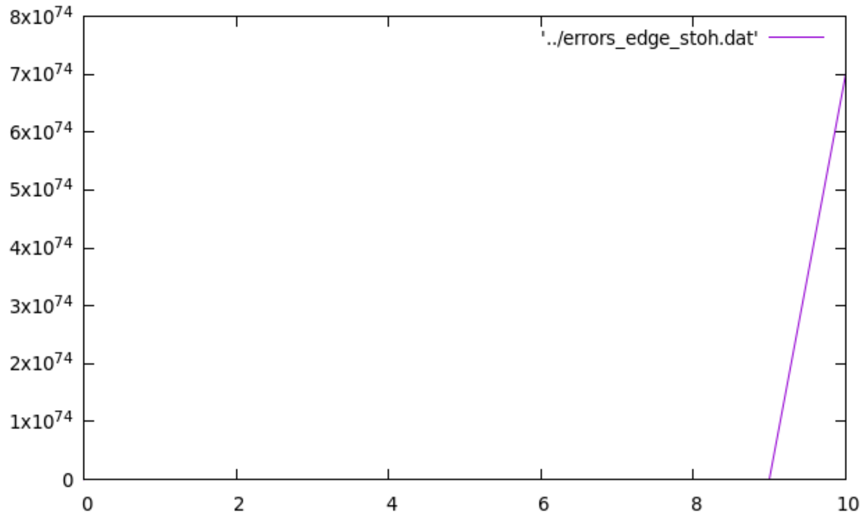

Figure 21: SGD Divergencija ukupne greške

Za premale stope učenja sustav jako sporo konvergira. Primjer: SGD:η_1 : 0.00000000001, η_2 : 0.000007, $error$: $error$: $6.158311311006211e + 51$

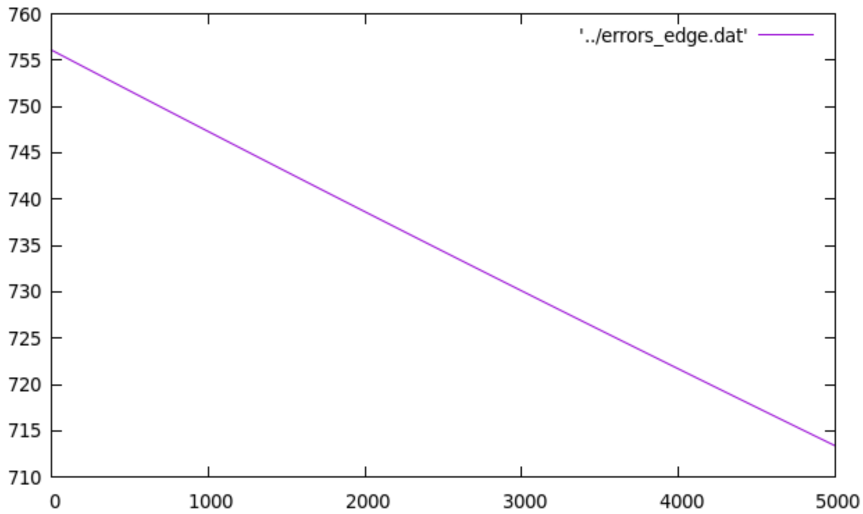

Figure 22: GD Prikaz konvergencije greške za jako male stope učenja

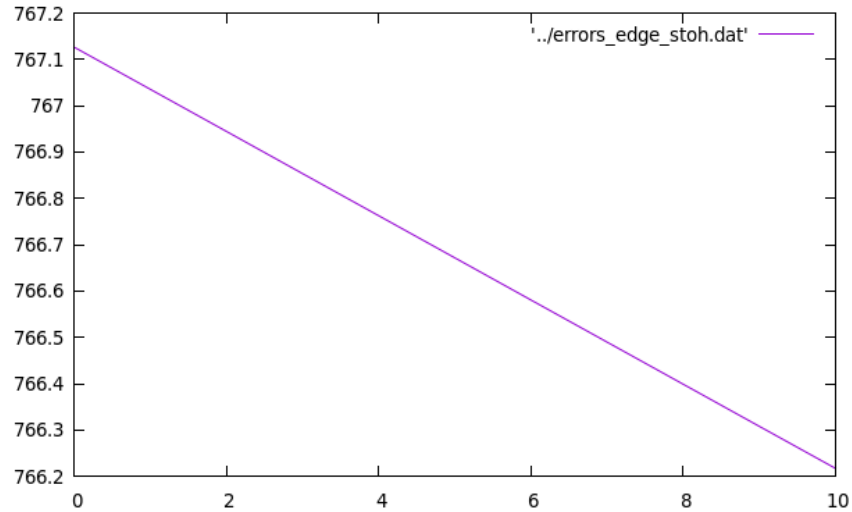

Figure 23: SGD Prikaz konvergencije greške za jako male stop učenja