

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA
NEIZRAZITO, EVOLUCIJSKO I NEURO RAČUNARSTVO
Ak. godina 2020./2021.

IZVJEŠTAJ

Šesta domaća zadaća - ANFIS

Jelena Bratulić

Voditelj: *doc.dr.sc. Marko Čupić*

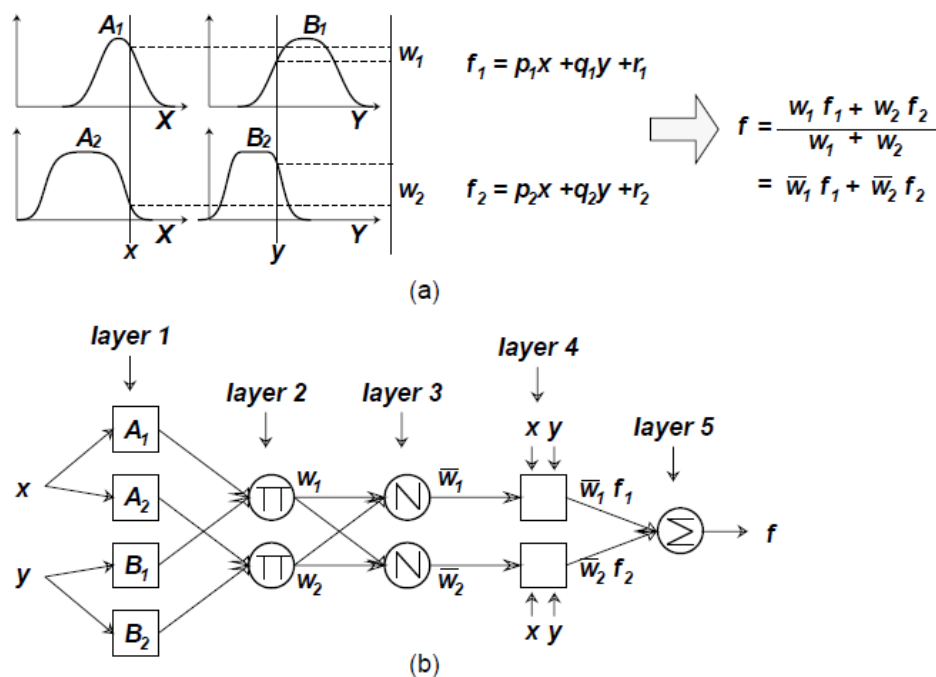
Zagreb, prosinac 2020.

SADRŽAJ

1. Izvod postupka učenja	1
2. Funkcija f	6
3. Učenje ANFIS mreže	8
4. Vizualizacija funkcije pripadnosti	11
5. Vizualizacija funkcije pogreške	13
6. Usporedba algoritama učenja (Batch vs. SGD)	15
7. Utjecaj stope učenja η	16

1. Izvod postupka učenja

Kako bismo mogli izvesti postupak učenja za ANFIS mrežu uz zaključivanje tipa 3 (TSK), potrebno je definirati parametre modela. Sam model prikazan je na 1.1. Primijetimo kako će naš model prilikom učenja optimizirati čak 7 različitih parametara: $a_i, b_i, c_i, d_i, p_i, q_i, r_i$ gdje će se indeks i kretati poprimati vrijednosti od 0 do m - broj pravila sustava. Također, sami ulazi u sustav bit će prikazani kao x i y , a izlaz iz sustava, odnosno predikcija modela kao z .



Slika 1.1: ANFIS mreža koja je implementirana u domaćoj zadaći.

Definirajmo prijenosne funkcije neurona iz prvog sloja:

$$\alpha_i = A_i = \frac{1}{1 + e^{b_i(x-a_i)}}, \beta_i = B_i = \frac{1}{1 + e^{d_i(y-c_i)}} \quad (1.1)$$

Korištenjem produkta kao T-norme, izlaz iz drugog sloja možemo zapisati kao:

$$\pi_i = \alpha_i \cdot \beta_i \quad (1.2)$$

Sama funkcija koju definira TSK zaključivanje definiramo kao:

$$z_i = p_i \cdot x + q_i \cdot y + r_i \quad (1.3)$$

Izlaz iz mreže, odnosno predikcija je tada određena kao:

$$o = \frac{\sum \pi_i \cdot z_i}{\sum \pi_i} \quad (1.4)$$

Funkcija MSE koju učenjem nastojimo minimizirati po definiciji je:

$$E_k = \frac{1}{2}(z_k - o_k)^2, E = \sum E_k \quad (1.5)$$

Primijetimo kako ovisno o vrsti algoritma učenja, koristit ćemo ili E_k ili E pri čemu će se E_k koristiti kod *on-line* verzije, a E kod *batch* verzije učenja. Sam izvod formula provest će se za *on-line* verziju učenja, no razlika u odnosu na *batch* verziju jest u tome što će se kod *online* učenja ažuriranje parametara odvijati nakon svakog primjera, a kod *batch* učenja će se gradijent izračunati za sve primjere iz skupa podataka i tek onda će se provesti ažuriranje parametara.

Kako bismo proveli postupak učenja, potrebno je izraziti parcijalne derivacije funkcije MSE za parametre modela kao:

$$\frac{\partial E_k}{\partial a_i} = \frac{\partial E_k}{\partial o_k} \frac{\partial o_k}{\partial \pi_i} \frac{\partial \pi_i}{\partial \alpha_i} \frac{\partial \alpha_i}{\partial a_i} \quad (1.6)$$

$$\frac{\partial E_k}{\partial b_i} = \frac{\partial E_k}{\partial o_k} \frac{\partial o_k}{\partial \pi_i} \frac{\partial \pi_i}{\partial \alpha_i} \frac{\partial \alpha_i}{\partial b_i} \quad (1.7)$$

$$\frac{\partial E_k}{\partial c_i} = \frac{\partial E_k}{\partial o_k} \frac{\partial o_k}{\partial \pi_i} \frac{\partial \pi_i}{\partial \beta_i} \frac{\partial \beta_i}{\partial c_i} \quad (1.8)$$

$$\frac{\partial E_k}{\partial d_i} = \frac{\partial E_k}{\partial o_k} \frac{\partial o_k}{\partial \pi_i} \frac{\partial \pi_i}{\partial \beta_i} \frac{\partial \beta_i}{\partial d_i} \quad (1.9)$$

$$\frac{\partial E_k}{\partial p_i} = \frac{\partial E_k}{\partial o_k} \frac{\partial o_k}{\partial z_i} \frac{\partial z_i}{\partial p_i} \quad (1.10)$$

$$\frac{\partial E_k}{\partial q_i} = \frac{\partial E_k}{\partial o_k} \frac{\partial o_k}{\partial z_i} \frac{\partial z_i}{\partial q_i} \quad (1.11)$$

$$\frac{\partial E_k}{\partial r_i} = \frac{\partial E_k}{\partial o_k} \frac{\partial o_k}{\partial z_i} \frac{\partial z_i}{\partial r_i} \quad (1.12)$$

Raspišimo redom parcijalne derivacije:

$$\frac{\partial E_k}{\partial o_k} = -(z_k - o_k) \quad (1.13)$$

$$\frac{\partial o_k}{\partial \pi_i} = \frac{\partial}{\partial \pi_i} \frac{\sum \pi_i \cdot z_i}{\sum \pi_i} = \frac{\sum_{j=1, j \neq i}^m \pi_j \cdot (z_i - z_j)}{(\sum_{j=1}^m \pi_j)^2} \quad (1.14)$$

$$\frac{\partial \pi_i}{\partial \alpha_i} = \beta_i, \quad \frac{\partial \pi_i}{\partial \beta_i} = \alpha_i \quad (1.15)$$

$$\begin{aligned} \frac{\partial \alpha_i}{\partial a_i} &= - \left[\frac{1}{1 + e^{b_i(x-a_i)}} \right]^2 \cdot \frac{\partial}{\partial a_i} (1 + e^{b_i(x-a_i)}) = \left[\frac{1}{1 + e^{b_i(x-a_i)}} \right]^2 \cdot b_i \cdot e^{b_i(x-a_i)} \\ &= b_i \cdot \frac{e^{b_i(x-a_i)}}{(1 + e^{b_i(x-a_i)})^2} = b_i \cdot \frac{1 - 1 + e^{b_i(x-a_i)}}{(1 + e^{b_i(x-a_i)})^2} = b_i \cdot \frac{-1}{(1 + e^{b_i(x-a_i)})^2} + \frac{e^{b_i(x-a_i)} + 1}{(1 + e^{b_i(x-a_i)})^2} \\ &= b_i \cdot \frac{1}{1 + e^{b_i(x-a_i)}} \cdot \left(1 - \frac{1}{1 + e^{b_i(x-a_i)}} \right) = b_i \cdot \alpha_i \cdot (1 - \alpha_i) \end{aligned} \quad (1.16)$$

$$\begin{aligned} \frac{\partial \alpha_i}{\partial b_i} &= - \left[\frac{1}{1 + e^{b_i(x-a_i)}} \right]^2 \cdot \frac{\partial}{\partial b_i} (1 + e^{b_i(x-a_i)}) = -(x - a_i) \cdot \left[\frac{1}{1 + e^{b_i(x-a_i)}} \right]^2 \cdot e^{b_i(x-a_i)} \\ &= -(x - a_i) \cdot \frac{e^{b_i(x-a_i)}}{(1 + e^{b_i(x-a_i)})^2} = -(x - a_i) \cdot \frac{1 - 1 + e^{b_i(x-a_i)}}{(1 + e^{b_i(x-a_i)})^2} \\ &= -(x - a_i) \cdot \frac{-1}{(1 + e^{b_i(x-a_i)})^2} + \frac{e^{b_i(x-a_i)} + 1}{(1 + e^{b_i(x-a_i)})^2} \\ &= -(x - a_i) \cdot \frac{1}{1 + e^{b_i(x-a_i)}} \cdot \left(1 - \frac{1}{1 + e^{b_i(x-a_i)}} \right) = -(x - a_i) \cdot \alpha_i \cdot (1 - \alpha_i) \end{aligned} \quad (1.17)$$

$$\begin{aligned} \frac{\partial \beta_i}{\partial c_i} &= - \left[\frac{1}{1 + e^{d_i(y-c_i)}} \right]^2 \cdot \frac{\partial}{\partial c_i} (1 + e^{d_i(y-c_i)}) = \left[\frac{1}{1 + e^{d_i(y-c_i)}} \right]^2 \cdot d_i \cdot e^{d_i(y-c_i)} \\ &= d_i \cdot \frac{e^{d_i(y-c_i)}}{(1 + e^{d_i(y-c_i)})^2} = d_i \cdot \frac{1 - 1 + e^{d_i(y-c_i)}}{(1 + e^{d_i(y-c_i)})^2} = d_i \cdot \frac{-1}{(1 + e^{d_i(y-c_i)})^2} + \frac{e^{d_i(y-c_i)} + 1}{(1 + e^{d_i(y-c_i)})^2} \\ &= d_i \cdot \frac{1}{1 + e^{d_i(y-c_i)}} \cdot \left(1 - \frac{1}{1 + e^{d_i(y-c_i)}} \right) = d_i \cdot \beta_i \cdot (1 - \beta_i) \end{aligned} \quad (1.18)$$

$$\begin{aligned}
\frac{\partial \beta_i}{\partial d_i} &= - \left[\frac{1}{1 + e^{d_i(y-c_i)}} \right]^2 \cdot \frac{\partial}{\partial d_i} (1 + e^{d_i(y-c_i)}) = -(y - c_i) \cdot \left[\frac{1}{1 + e^{d_i(y-c_i)}} \right]^2 \cdot e^{d_i(y-c_i)} \\
&= -(y - c_i) \cdot \frac{e^{d_i(y-c_i)}}{(1 + e^{d_i(y-c_i)})^2} = -(y - c_i) \cdot \frac{1 - 1 + e^{d_i(y-c_i)}}{(1 + e^{d_i(y-c_i)})^2} \\
&= -(y - c_i) \cdot \frac{-1}{(1 + e^{d_i(y-c_i)})^2} + \frac{e^{d_i(y-c_i)} + 1}{(1 + e^{d_i(y-c_i)})^2} \\
&= -(y - c_i) \cdot \frac{1}{1 + e^{d_i(y-c_i)}} \cdot \left(1 - \frac{1}{1 + e^{d_i(y-c_i)}} \right) = -(y - c_i) \cdot \beta_i \cdot (1 - \beta_i) \quad (1.19)
\end{aligned}$$

$$\frac{\partial o_k}{\partial z_i} = \frac{\partial}{\partial \pi_i} \frac{\sum \pi_i \cdot z_i}{\sum \pi_i} = \frac{\pi_i}{\sum_{j=1}^m \pi_j} \quad (1.20)$$

$$\frac{\partial z_i}{\partial p_i} = x, \quad \frac{\partial z_i}{\partial q_i} = y, \quad \frac{\partial z_i}{\partial r_i} = 1 \quad (1.21)$$

Konačno, kompletne parcijalne derivacije dobijemo množenjem pojedinačnih derivacija:

$$\frac{\partial E_k}{\partial a_i} = -(z_k - o_k) \cdot \frac{\sum_{j=1, j \neq i}^m \pi_j \cdot (z_i - z_j)}{(\sum_{j=1}^m \pi_j)^2} \cdot \beta_i \cdot b_i \cdot \alpha_i \cdot (1 - \alpha_i) \quad (1.22)$$

$$\frac{\partial E_k}{\partial b_i} = (z_k - o_k) \cdot \frac{\sum_{j=1, j \neq i}^m \pi_j \cdot (z_i - z_j)}{(\sum_{j=1}^m \pi_j)^2} \cdot \beta_i \cdot (x - a_i) \cdot \alpha_i \cdot (1 - \alpha_i) \quad (1.23)$$

$$\frac{\partial E_k}{\partial c_i} = -(z_k - o_k) \cdot \frac{\sum_{j=1, j \neq i}^m \pi_j \cdot (z_i - z_j)}{(\sum_{j=1}^m \pi_j)^2} \cdot \alpha_i \cdot d_i \cdot \beta_i \cdot (1 - \beta_i) \quad (1.24)$$

$$\frac{\partial E_k}{\partial d_i} = (z_k - o_k) \cdot \frac{\sum_{j=1, j \neq i}^m \pi_j \cdot (z_i - z_j)}{(\sum_{j=1}^m \pi_j)^2} \cdot \alpha_i \cdot (y - c_i) \cdot \beta_i \cdot (1 - \beta_i) \quad (1.25)$$

$$\frac{\partial E_k}{\partial p_i} = -(z_k - o_k) \cdot \frac{\pi_i}{\sum_{j=1}^m \pi_j} \cdot x \quad (1.26)$$

$$\frac{\partial E_k}{\partial q_i} = -(z_k - o_k) \cdot \frac{\pi_i}{\sum_{j=1}^m \pi_j} \cdot y \quad (1.27)$$

$$\frac{\partial E_k}{\partial r_i} = -(z_k - o_k) \cdot \frac{\pi_i}{\sum_{j=1}^m \pi_j} \quad (1.28)$$

Ažuriranje parametara provest ćemo po formuli koju smo već uveli kod Backpropagation algoritma:

$$w^{k+1} = w^k - \phi \cdot \frac{\partial E_k}{\partial w} \quad (1.29)$$

što za naše parametre $a_i, b_i, c_i, d_i, p_i, q_i$ i r_i daje sljedeće izraze:

$$a_i^{t+1} = a_i^t + \eta \cdot (z_k - o_k) \cdot \frac{\sum_{j=1, j \neq i}^m \pi_j \cdot (z_i - z_j)}{(\sum_{j=1}^m \pi_j)^2} \cdot \beta_i \cdot b_i \cdot \alpha_i \cdot (1 - \alpha_i) \quad (1.30)$$

$$b_i^{t+1} = b_i^t - \eta \cdot (z_k - o_k) \cdot \frac{\sum_{j=1, j \neq i}^m \pi_j \cdot (z_i - z_j)}{(\sum_{j=1}^m \pi_j)^2} \cdot \beta_i \cdot (x - a_i) \cdot \alpha_i \cdot (1 - \alpha_i) \quad (1.31)$$

$$c_i^{t+1} = c_i^t + \eta \cdot (z_k - o_k) \cdot \frac{\sum_{j=1, j \neq i}^m \pi_j \cdot (z_i - z_j)}{(\sum_{j=1}^m \pi_j)^2} \cdot \alpha_i \cdot d_i \cdot \beta_i \cdot (1 - \beta_i) \quad (1.32)$$

$$d_i^{t+1} = d_i^t - \eta \cdot (z_k - o_k) \cdot \frac{\sum_{j=1, j \neq i}^m \pi_j \cdot (z_i - z_j)}{(\sum_{j=1}^m \pi_j)^2} \cdot \alpha_i \cdot (y - c_i) \cdot \beta_i \cdot (1 - \beta_i) \quad (1.33)$$

$$p_i^{t+1} = p_i^t + \eta \cdot (z_k - o_k) \cdot \frac{\pi_i}{\sum_{j=1}^m \pi_j} \cdot x \quad (1.34)$$

$$q_i^{t+1} = q_i^t + \eta \cdot (z_k - o_k) \cdot \frac{\pi_i}{\sum_{j=1}^m \pi_j} \cdot y \quad (1.35)$$

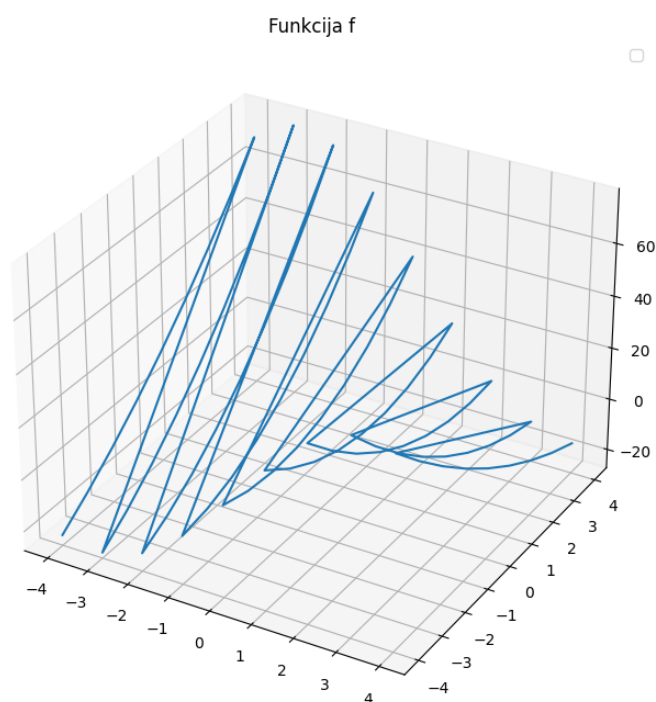
$$r_i^{t+1} = r_i^t + \eta \cdot (z_k - o_k) \cdot \frac{\pi_i}{\sum_{j=1}^m \pi_j} \quad (1.36)$$

2. Funkcija f

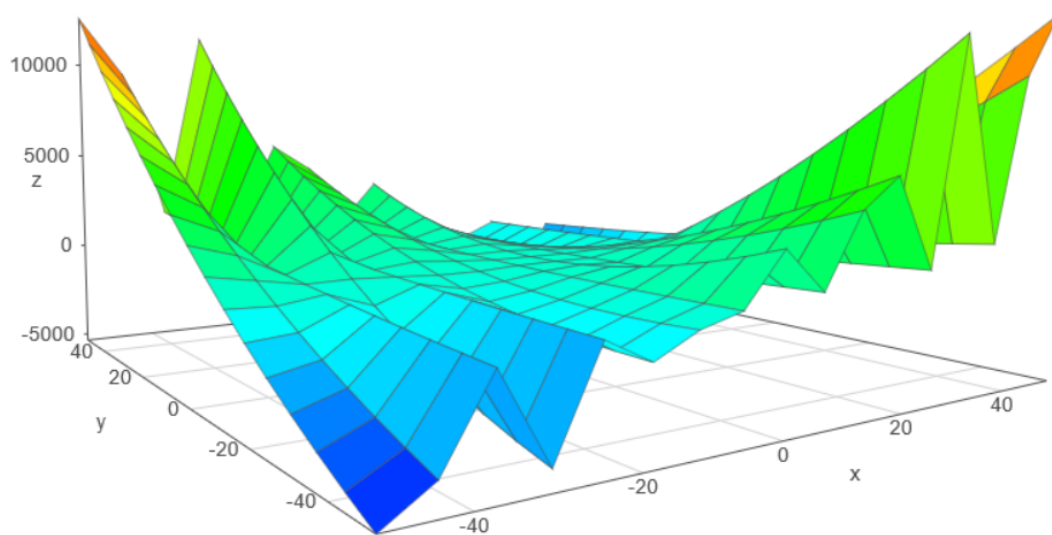
Model mreže ANFIS mora naučiti potrebne parametre kako bi mogao aproksimirati funkciju

$$f = ((x - 1)^2 + (y + 2)^2 - 5xy + 3) \cdot \cos^2\left(\frac{x}{5}\right) \quad (2.1)$$

Ulazi x, y bit će u rasponu $[-4, 4]$ te će se konačni skup podataka formirati kao cjelobrojne kombinacije ulaza x i y za koje će se izračunati pripadna vrijednost funkcije f (z koordinata). Na sljedećem grafu je prikazana funkcija f za prethodno spomenuti raspon vrijednosti čime je ujedno prikazana i očekivan izlaz iz mreže.



Slika 2.1: Funkcija f za dani raspon brojeva x i y .



Slika 2.2: Funkcija f na široj domeni i prikazana kao ploha.

3. Učenje ANFIS mreže

Samo učenje mreže isprobano je s jednim, dva, pet te deset pravila. Naravno, s obzirom na društvi broj pravila, dobiveno je i drukčije ponašanje mreže koje se može sažeti na sljedeći način. Koristeći više pravila, povećalo se vrijeme izvođenja jedne epohe, a time i ukupno izvođenje jednog treninga.

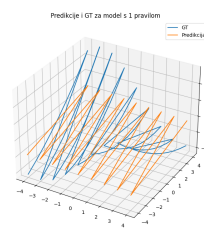
Jedno pravilo nije bilo dovoljno kako bi ANFIS mogla dobro generalizirati. Tek uz malu stopu učenja od $\eta = 1 \cdot 10^{-7}$ postignuta je konvergencija prema minimumu, ali je iznimno sporo konvergirala, odnosno u 100000 epoha greška se smanjila tek na 253. Smanjenje MSE nije bilo uvelike različito ukoliko se koristio potpuni gradijent ili aproksimirani gradijent. Koristeći stopu učenja veću od spomenute, model bi zaglavio u lokalnom minimumu i ne bi se uspio pomaknuti iz njega. Također, valja spomenuti kako su zapravo korištene različite vrijednosti stopa učenja za ažuriranje parametara prijenosih funkcija u odnosu na parametre funkcije z. U pravilu je stopa učenja za parametre prijenosne funkcije bila 100 puta manja naspram one stope učenja koja je spomenuta.

Dva pravila poboljšala su generalizaciju same mreže, ali se nastavio problem zapinjanja u lokalnim minimumima. Uz stopu učenja od $\eta = 1 \cdot 10^{-4}$ postignuti su optimalni rezultati i greška MSE se s obzirom na prethodni slučaj uvelike smanjila. Uz korištenje spomenute stope učenja, greška se smanjila na 3. Zanimljivo je kako su bolji rezultati dobiveni uz korištenje *batch* verzije algoritma učenja, dok je uz korištenje *online* verzije greška MSE bila 10 puta veća.

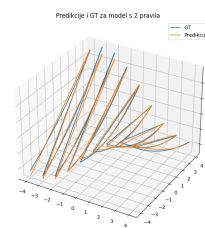
Pet pravila dodatno su poboljšala ponašanje mreže ANFIS. Optimalni rezultati dobiveni su uz stopu učenja od $\eta = 1 \cdot 10^{-4}$. Uz korištenje *batch* učenja, greška MSE se spustila na 0.29, dok je uz *online* učenje greška MSE bila 0.15.

Deset pravila dala su najbolje rezultate te je upravo verzija modela s 10 pravila korištena u daljnjim eksperimentima i vizualizacijama. Optimalni rezultati postignuti su uz stopu učenja od $\eta = 1 \cdot 10^{-3}$ gdje je dobivena MSE od 0.0026 za *batch* učenje te 0.0069 za *online* učenje. Korištenjem manjih stopa učenja je učenje sporije konvergiralo i u istom broju epoha je smanjenje MSE bilo manje, dok se za veće stope učenja ponovno pojavio problem lokalnih minimuma.

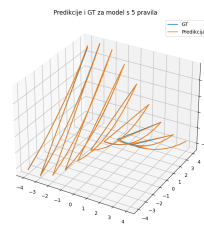
Na sljedećim grafovima su prikazani izlazi dobiveni iz modela za određeni broj pravila. Također je prikazana i greška modela koja je definirana kao $\delta(z) = z_{predikcija} - z_{istina}$.



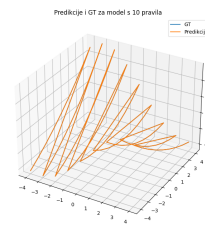
(a) Jedno pravilo.



(b) Dva pravila.

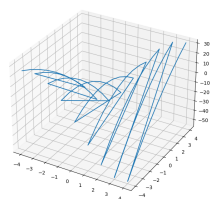


(c) Pet pravila.

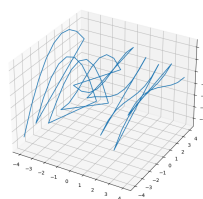


(d) Deset pravila.

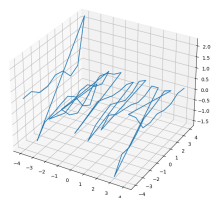
Slika 3.1: Usporedba predikcija i očekivanog izlaza iz modela za različit broj pravila.



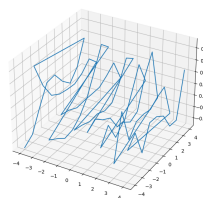
(a) Jedno pravilo.



(b) Dva pravila.



(c) Pet pravila.

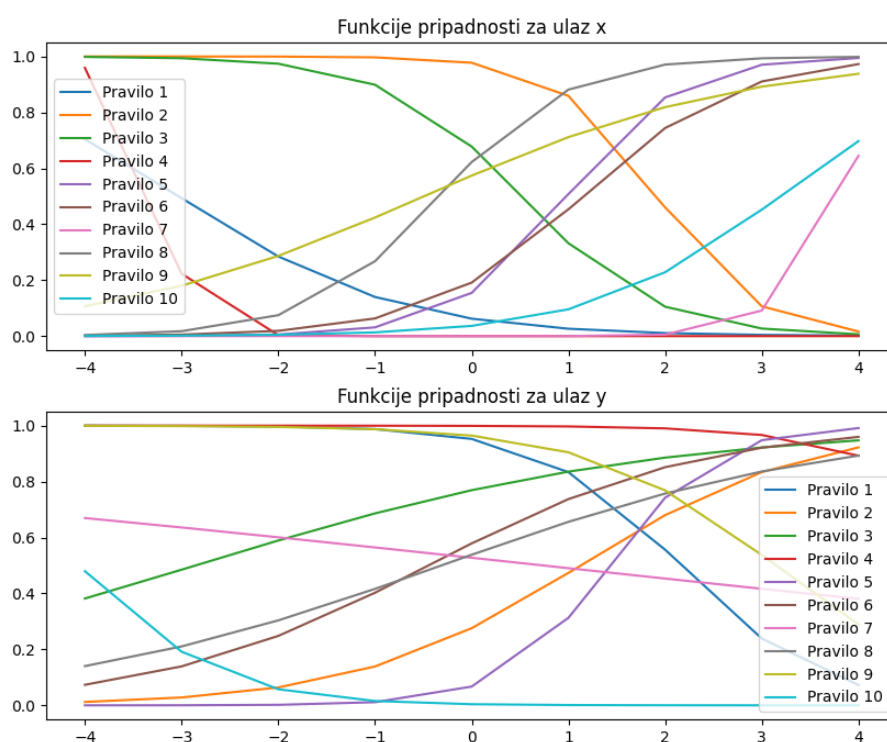


(d) Deset pravila.

Slika 3.2: Funkcija pogreške za modele s različitim brojem parametara.

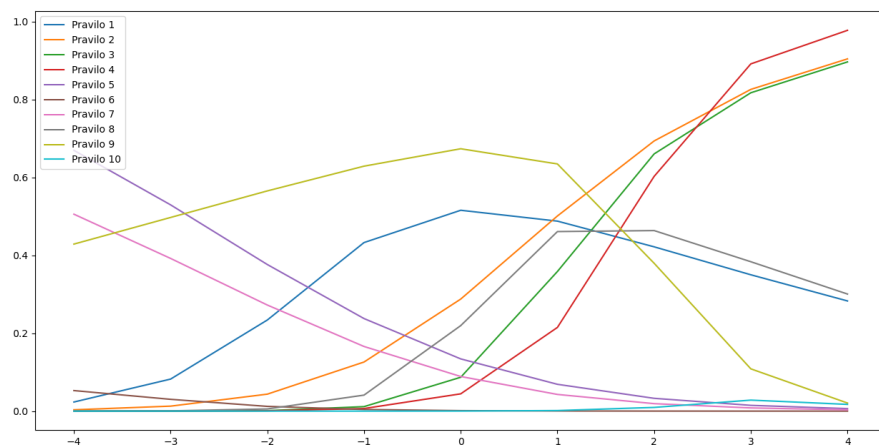
4. Vizualizacija funkcije pripadnosti

Naučene parametre funkcija pripadnosti prikazat ću za model s najboljim ponašanjem, dakle model s 10 pravila.



Slika 4.1: Funkcije pripadnosti za naučene parametre a, b, c, d .

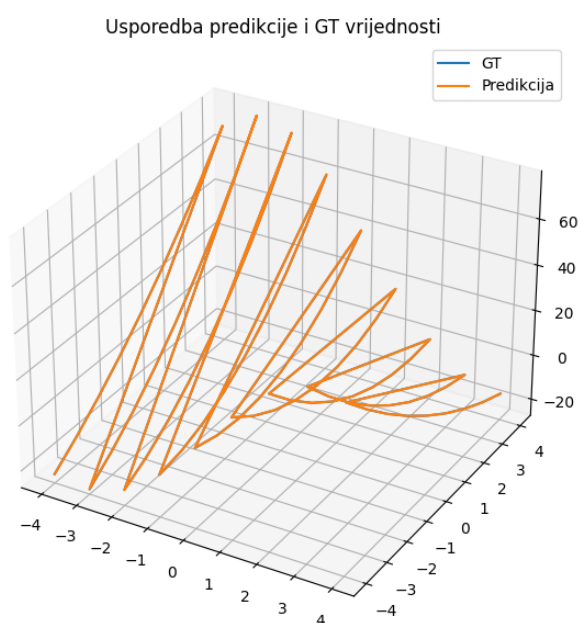
Možemo primijetiti kako je prijenosna funkcija jednog pravila za ulaz x 0, a za ulaz y ona je pravac. U nastavku su još vizualizirane i vrijednosti parametra π_i koje su dobivene kao umnožak prijenosnih funkcija za izlaz x i za izlaz y .



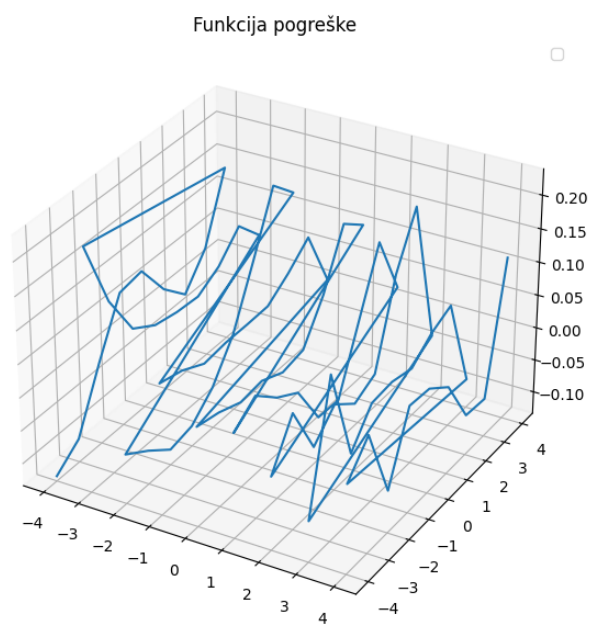
Slika 4.2: Vizualizacija funkcije pripadnosti preko parametra π .

5. Vizualizacija funkcije pogreške

Na sljedećim grafovima prikazane su dobivene predikcije vrijednosti z iz modela te stvarne vrijednosti funkcije f za dane x, y u rasponu od -4 do 4. Također je prikazana i greška samog modela kao $\delta(x) = z_{predikcija} - z_{istinito}$.



Slika 5.1: Usporedba dobivenih i očekivanih predikcija za funkciju f .



Slika 5.2: Pogreška modela

6. Usporedba algoritama učenja (Batch vs. SGD)

Usporedba algoritma učenja prikazana je na sljedećem grafu. Usporedba je napravljena za model od 10 primjera i sa stopama učenja od $\eta_1 = 5 \cdot 10^{-4}$ i $\eta_2 = 1 \cdot 10^{-3}$ pri čemu je η_1 stopa učenja za ažuriranje parametara a, b, c, d , a η_2 stopa učenja za ažuriranje parametara p, q, r .



Slika 6.1: Usporedba algoritama učenja prikazana nad svih 10000 epoha učenja

Promatrajući graf vidimo kako nema velike razlike u korištenju *batch* ili *on-line* verzije učenja. Razlika je ipak vidljiva u krajnjim vrijednostima funkcije gubitka MSE koja je kod *batch* verzija niža u odnosu na onu dobivenu s *on-line* verzijom. Također, kod *on-line* verzije vidimo kako postoje povremeni skokovi vrijednosti MSE.

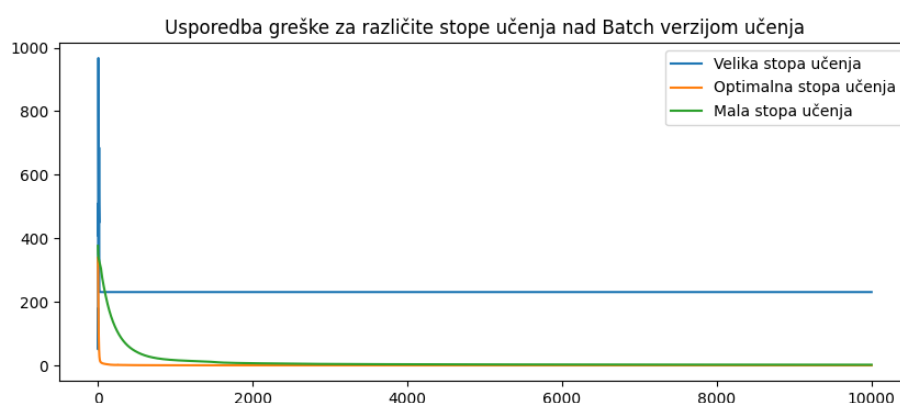
7. Utjecaj stope učenja η

Utjecaj različitih stopa učenja prikazat ću nad modelom s 10 pravila učenja. Za usporedbu su odabrane 3 kombinacije stopa učenja pri čemu je η_1 stopa učenja za ažuriranje parametara a, b, c, d , a η_2 stopa učenja za ažuriranje parametara p, q, r .

Oznaka	η_1	η_2
Velika stopa učenja	$5 \cdot 10^{-3}$	$1 \cdot 10^{-2}$
Optimalna stopa učenja	$5 \cdot 10^{-4}$	$1 \cdot 10^{-3}$
Mala stopa učenja	$5 \cdot 10^{-4}$	$1 \cdot 10^{-5}$

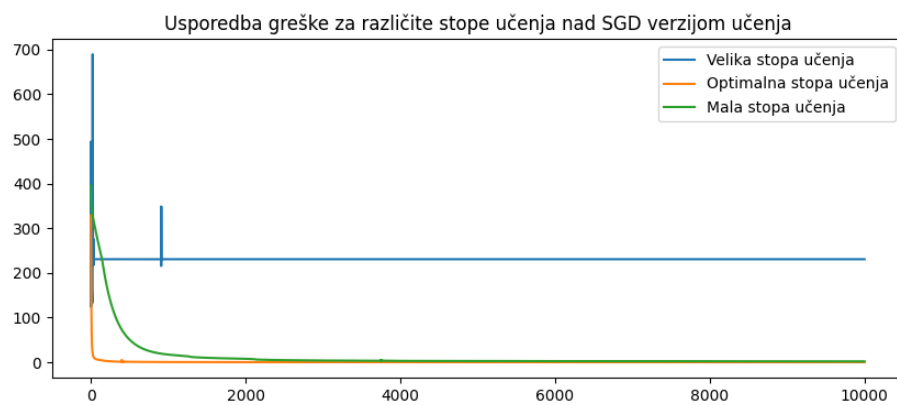
Tablica 7.1: Stope učenja korištenje za usporedbu uspješnosti modela.

Na sljedećim grafovima prikazano je kretanje funkcije gubitka MSE kroz epohe zasebno za *batch* i *on-line* verzije učenja.



Slika 7.1: Usporedba različitih stopa učenja nad Batch verzijom učenja

Analizirajući grafove za obje verzije algoritma učenja vidimo da velika stopa učenja uzrokuje zaglavljivanje u lokalnom minimumu, dok manja stopa učenja uzrokuje sporiju konvergenciju od one kod optimalne stope učenja.



Slika 7.2: Usporedba različitih stopa učenja nad SGD verzijom učenja