1 Prvi sloj i njegove parcijalne derivacije

Ovdje su izvodi svih parcijalnih derivacija izlaza prvog sloja po parametrima. To će nam trebati kasnije kada ćemo izvoditi učenje svakog od parametara.

$$A_i = \frac{1}{1+e^{b_i(x-a_i)}}$$
$$\frac{\partial A_i}{\partial a_i} = \frac{1}{(1+e^{b_i(x-a_i)})^2}e^{b_i(x-a_i)}b_i = A_i^2(\frac{1}{A_i}-1)b_i = A_i(1-A_i)b_i$$
$$\frac{\partial A_i}{\partial b_i} = \frac{-1}{(1+e^{b_i(x-a_i)})^2}e^{b_i(x-a_i)}(x-a_i) = -A_i^2(\frac{1}{A_i}-1)(x-a_i) = -A_i(1-A_i)(x-a_i)$$

$$B_i = \frac{1}{1+e^{d_i(y-c_i)}}$$
$$\frac{\partial B_i}{\partial c_i} = \frac{1}{(1+e^{d_i(x-c_i)})^2}e^{d_i(y-c_i)}d_i = B_i^2(\frac{1}{B_i}-1)d_i = B_i(1-B_i)d_i$$
$$\frac{\partial B_i}{\partial d_i} = \frac{-1}{(1+e^{d_i(x-c_i)})^2}e^{d_i(y-c_i)}(y-c_i) = -B_i^2(\frac{1}{B_i}-1)(y-c_i) = -B_i(1-B_i)(y-c_i)$$

2 Drugi sloj i njegove parcijalne derivacije

$$w_i = A_i B_i$$

$$\frac{\partial w_i}{\partial A_i} = B_i \qquad \frac{\partial w_i}{\partial B_i} = A_i$$

3 Treći sloj i njegove parcijalne derivacije

Ovaj sloj je normalizacijski sloj za kojeg neću posebno izvoditi derivacije, jer je lakše koristiti samo tu sumu kasnije u računu. Normalizacija svake od dobivenih težina provodi se na sljedeći način:

$$\overline{w}_i = \frac{w_i}{\sum_{j=1}^{M} w_j}$$

4 Četvrti i peti sloj i njihove parcijalne derivacije

$$f_i = p_i x + q_i y + r_i$$
$$f = \frac{\sum_{i=1}^{M} w_i f_i}{\sum_{i=1}^{M} w_i}$$

$$\frac{\partial f}{\partial f_i} = \frac{w_i}{\sum_{i=1}^{M} w_i} = \overline{w}_i \qquad \frac{\partial f}{\partial w_i} = \frac{f_i \sum_{j=1}^{M} w_j - \sum_{j=1}^{M} f_j w_j}{(\sum_{j=1}^{M} w_j)^2} = \frac{\sum_{j=1}^{M} w_j (f_i - f_j)}{(\sum_{j=1}^{M} w_j)^2}$$

$$\frac{\partial f_i}{\partial p_i} = x \qquad \frac{\partial f_i}{\partial q_i} = y \qquad \frac{\partial f_i}{\partial r_i} = 1$$

5 Pogreška izlaza mreže

E je oznaka pogreške za 1 podatak pri čemu je y točna oznaka za podatak, a f je izlaz mreže za taj isti podatak.

$$E = \frac{1}{2}(y - f)^2$$

$$\frac{\partial E}{\partial f} = -(y - f)$$

6 Određivanje pravila učenja za svaki od parametara

Ideja je svaki od parametara (koji god da bio) mijenjati tako da se pomičemo u negativnom smjeru gradijenta pogreške E po tom parametru za dani podatak. Za proizvoljan parametar θ to bi glasilo:

$$\theta_{t+1} = \theta_t - \eta \frac{\partial E}{\partial \theta}$$

Budući da je gradijent (u ovom slučaju parcijalna derivacija) smjer najvećeg rasta funkcije, pomicanjem u negativnom smjeru gradijenta idemo u smjeru najvećeg pada ciljne funkcije (u ovom slučaju funkcije pogreške E) i to je ono što nam odgovara jer želimo minimizirati pogrešku. Hiperparametar η je stopa učenja kojom reguliramo da koraci budu dovoljno maleni kako ne bi došlo do divergencije gradijentnog spusta.

Odredimo sada parcijalne derivacije pogreške po svakom od parametara:

$$\frac{\partial E}{\partial a_i} = \frac{\partial E}{\partial f}\frac{\partial f}{\partial w_i}\frac{\partial w_i}{\partial A_i}\frac{\partial A_i}{\partial a_i} \qquad \frac{\partial E}{\partial p_i} = \frac{\partial E}{\partial f}\frac{\partial f}{\partial f_i}\frac{\partial f_i}{\partial p_i}$$

$$\frac{\partial E}{\partial b_i} = \frac{\partial E}{\partial f}\frac{\partial f}{\partial w_i}\frac{\partial w_i}{\partial A_i}\frac{\partial A_i}{\partial b_i} \qquad \frac{\partial E}{\partial q_i} = \frac{\partial E}{\partial f}\frac{\partial f}{\partial f_i}\frac{\partial f_i}{\partial q_i}$$

$$\frac{\partial E}{\partial c_i} = \frac{\partial E}{\partial f}\frac{\partial f}{\partial w_i}\frac{\partial w_i}{\partial B_i}\frac{\partial B_i}{\partial c_i} \qquad \frac{\partial E}{\partial r_i} = \frac{\partial E}{\partial f}\frac{\partial f}{\partial f_i}\frac{\partial f_i}{\partial r_i}$$

$$\frac{\partial E}{\partial d_i} = \frac{\partial E}{\partial f}\frac{\partial f}{\partial w_i}\frac{\partial w_i}{\partial B_i}\frac{\partial B_i}{\partial d_i}$$

Uvrštavanjem prije dobivenih izraza u svaku od jednadžbi dobivamo:

$$\frac{\partial E}{\partial a_i} = -(y-f)\frac{\sum_{j=1}^{M} w_j(f_i - f_j)}{(\sum_{j=1}^{M} w_j)^2} B_i A_i (1-A_i) b_i$$

$$\frac{\partial E}{\partial b_i} = -(y-f)\frac{\sum_{j=1}^{M} w_j(f_i - f_j)}{(\sum_{j=1}^{M} w_j)^2} B_i (-A_i)(1-A_i)(x-a_i)$$

$$\frac{\partial E}{\partial c_i} = -(y-f)\frac{\sum_{j=1}^{M} w_j(f_i - f_j)}{(\sum_{j=1}^{M} w_j)^2} A_i B_i (1-B_i) d_i$$

$$\frac{\partial E}{\partial d_i} = -(y-f)\frac{\sum_{j=1}^{M} w_j(f_i - f_j)}{(\sum_{j=1}^{M} w_j)^2} A_i (-B_i)(1-B_i)(y-c_i)$$

$$\frac{\partial E}{\partial p_i} = -(y-f)\frac{w_i}{\sum_{j=1}^{M} w_j} x \qquad \frac{\partial E}{\partial q_i} = -(y-f)\frac{w_i}{\sum_{j=1}^{M} w_j} y \qquad \frac{\partial E}{\partial r_i} = -(y-f)\frac{w_i}{\sum_{j=1}^{M} w_j}$$

Prema pravilu učenja za proizvoljan parametar obnavljat ćemo vrijednosti svakog od parametara u algoritmu. Međutim javljaju se dvije inačice algoritma učenja: stohastička i grupna. Stohastička se inačica algoritma ponaša tako da

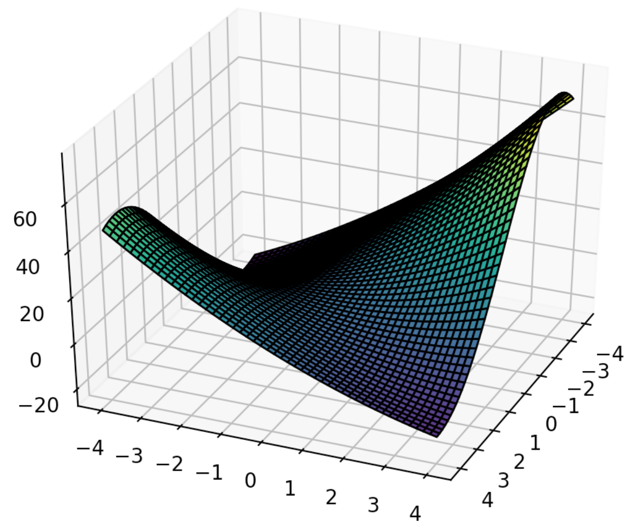

Abbildung 1: Ciljna funkcija

obnavlja parametere nakon svakog podatka koji prođe kroz mrežu, pa se tako u jednoj epohi parameteri obnavljaju N puta gdje je N broj podataka za učenje. Grupa se pak inačica ponaša tako da uprosječuje gradijente po svim uzorcima za učenje i tako se parameteri obnavljaju samo jednom u jednoj epohi učenja.

7 Ciljna funkcija

Ciljna funkcija glasi:

$$((x-1)^2 + (y+2)^2 - 5xy + 3) * cos^2(\frac{x}{5})$$

i njen izgled je prikazan na slici 1

8 Prikaz rada mreže za različiti broj pravila

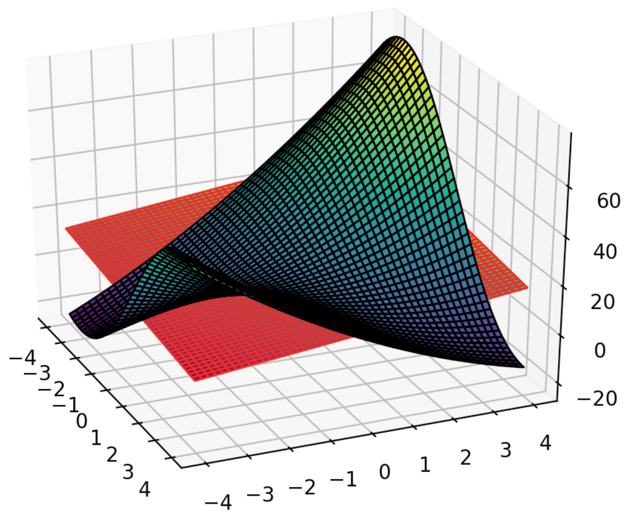

Abbildung 2: Izgled naučene funkcije za samo jedno pravilo.

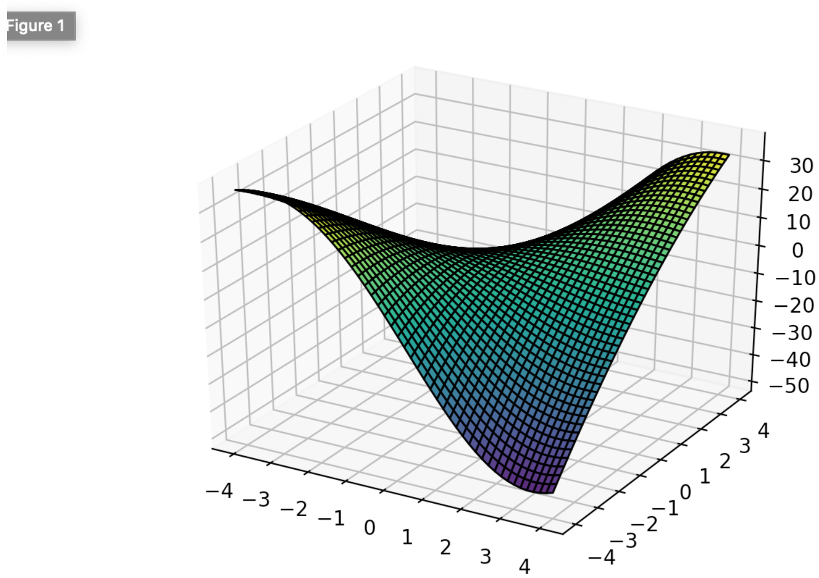

Abbildung 3: Ploha pogreške za samo jedno pravilo.

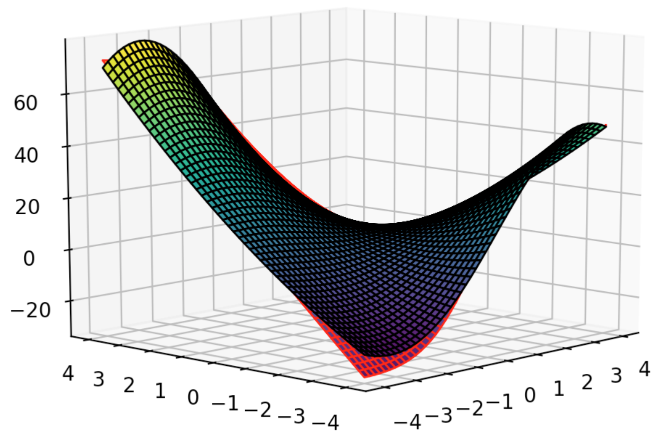

Abbildung 4: Izgled naučene funkcije za dva pravila.

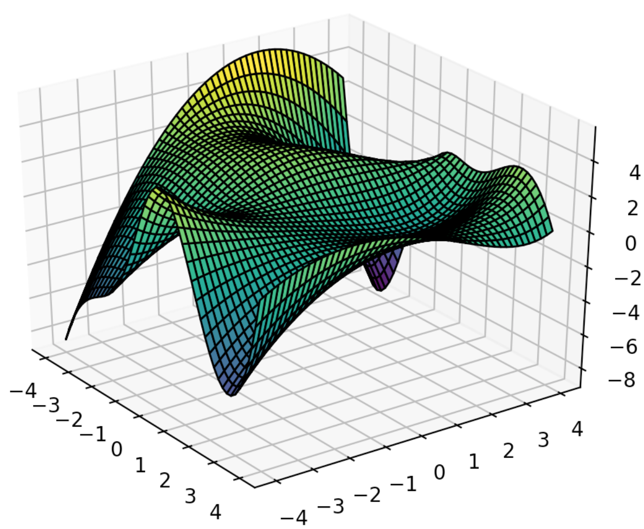

Abbildung 5: Ploha pogreške za dva pravila.

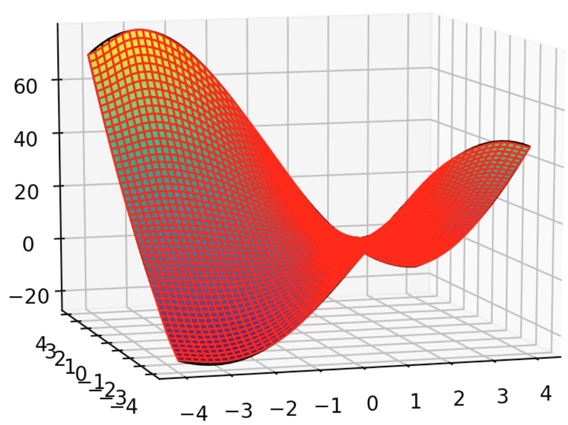

Abbildung 6: Izgled naučene funkcije za osam pravila.

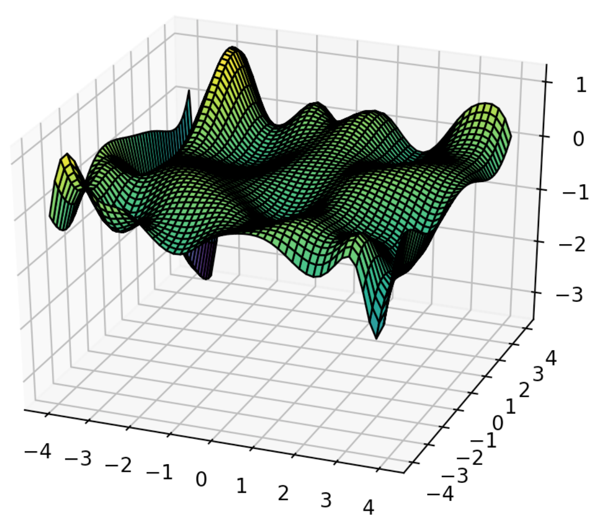

Abbildung 7: Ploha pogreške za osam pravila.

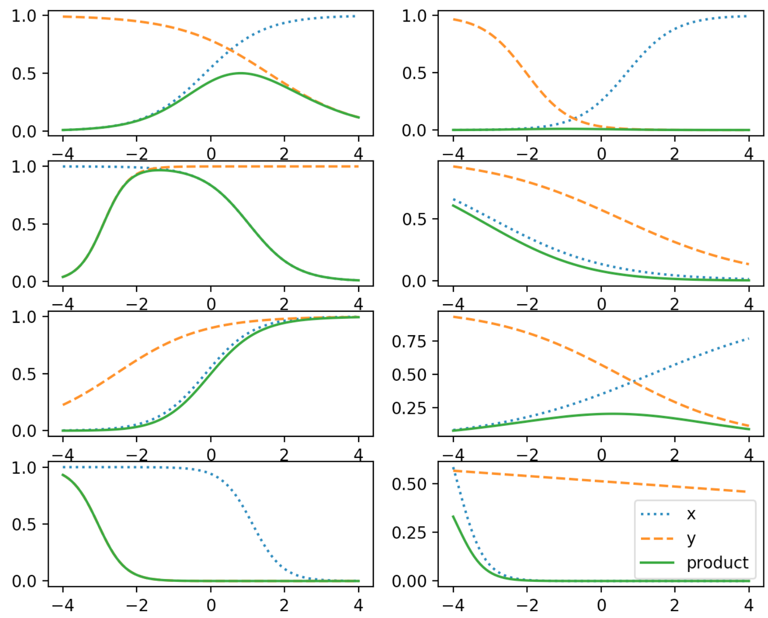

Abbildung 8: Naučene funkcije pripadnosti za 8 pravila.