

2. RBF neuronové sítě

Kapitola pojednává o neuronových sítích typu RBF. V kapitole je popsána základní struktura tohoto typu neuronové sítě. Poté následuje definice a charakteristika jednotlivých radiálně bazických funkcí, které slouží jako aktivační funkce neuronů. Důraz je kladen i na princip učení tohoto typu neuronové sítě, jež je odlišný od vícevrstvých perceptronových sítí.

2.1. Základní pojmy

Normou ($\|\cdot\|$) nazveme zobrazení z vektorového prostoru V na \mathbb{R} splňující:

- $\|x\| = 0 \Leftrightarrow x = 0$.
- Pro všechna $x, y \in V$ platí: $\|x+y\| \leq \|x\| + \|y\|$.
- Pro každé $x \in V$ a reálné číslo k platí, že $\|k \cdot x\| = |k| \cdot \|x\|$. [14]

Metrikou nazveme zobrazení ρ z množiny M^2 na \mathbb{R} :

- $\rho(x,y) = 0 \Leftrightarrow x = y$.
- Pro všechny $x, y \in M$ platí symetrie, tedy $\rho(x,y) = \rho(y,x)$.
- Pro všechny $x, y, z \in M$ platí trojúhelníková nerovnost, tedy $\rho(x,z) \leq \rho(x,y) + \rho(y,z)$. [14]

Metriku lze definovat pomocí normy $\rho(x,y) = \|x - y\|$. [14]

Graf G lze definovat jako uspořádanou dvojici (V,E) , kde V je neprázdná množina vrcholů (vertex) a E množina dvoubodových podmnožin reprezentující hrany (edge). Pokud je graf orientovaný, potom množina E obsahuje uspořádané dvojice ve tvaru (x,y) . Hrana $e = (x,y)$ začíná v uzlu x a končí v uzlu y . [17]

Neuronovou síť lze definovat jako orientovaný graf $G=(V,E)$, kde množina vrcholů (V) reprezentuje neurony a množina hran (E) jejich vzájemné propojení nazývané synapse. Množina V je rozdělena na disjunktní podmnožiny reprezentující jednotlivé vrstvy. Základním typem neuronové sítě je vícevrstvá perceptronová síť (Multi-Layer Perceptron, MLP). [15][18]

Učení neuronové sítě je proces adaptace synaptických vah. Při učení s učitelem je pro vstupní data znám optimální výstup. Ten je porovnán s výstupem neuronové sítě a pomocí zpětné vazby je provedena úprava synaptických vah.[11]

2.2. Radiálně bazické funkce

Radiálně bazické funkce jsou spojovány s teorií vícerozměrné interpolace. Obecně jsou definovány vztahem (2.1) kde ϕ reprezentuje použitou, zpravidla nelineární funkci, c je takzvané centrum neuronu a R metrika sloužící k určení vzdálenosti mezi vstupem x a centrem c . Obecně může být R definováno libovolnou normou $\|\cdot\|$ [11][22]

$$h(x) = \phi((x - c)^T R^{-1}(x - c)). \quad (2.1)$$

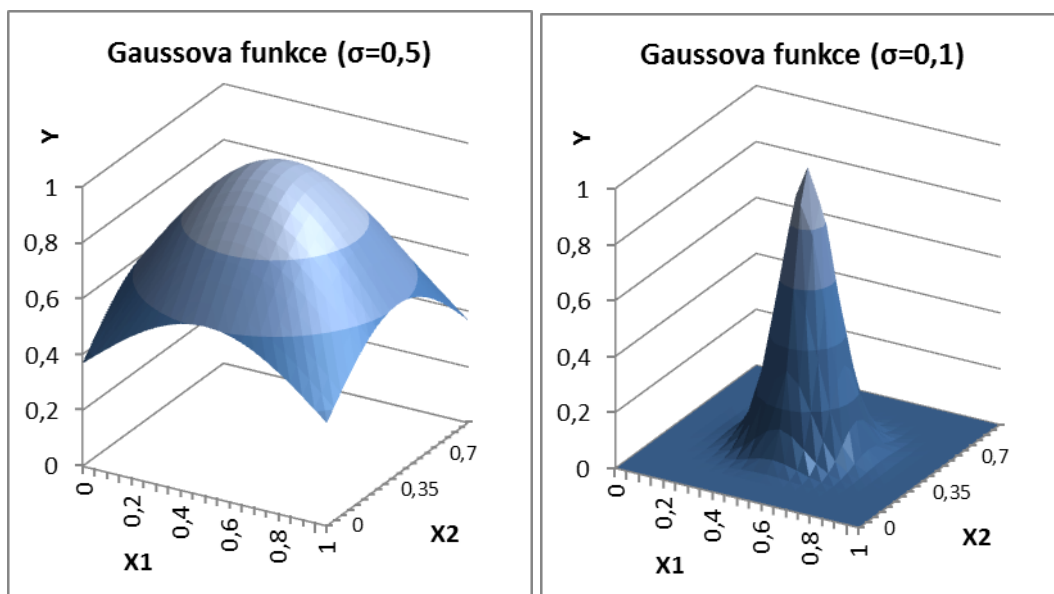
Níže jsou popsány tři nejčastěji používané radiálně bazické funkce. Symbol r značí metriku $\|x - c\|$. Nejčastěji používanou metrikou je Euklidovskou vzdálenost definovaná vztahem (2.2) [11][14]

$$\rho(\bar{x}, \bar{y}) = \|\bar{x} - \bar{y}\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}. \quad (2.2)$$

Gaussova funkce je definována rovnicí ve tvaru

$$\varphi(r) = \exp\left(-\frac{r^2}{2\sigma^2}\right) \text{ pro } \sigma > 0, r \in \mathbb{R}. \quad (2.3)$$

Na obr. 2 je zobrazen průběh Gaussovy funkce pro dvě proměnné x_1 a x_2 se středy v hodnotě 0,5 a poloměrem $\sigma = 0,5$ a $0,1$. Zatímco střed určuje pozici v prostoru, poloměr σ ovlivňuje tvar funkce. Určuje také hraniční vzdálenost, od které budou hodnoty funkce blízké či rovny nule. Čím blíže je vstupní vektor X středu, tím vyšší bude hodnota výstupní proměnné ϕ . Gaussova funkce je pro rostoucí hodnotu r , tedy vzdálenost od středu, klesající do nuly. Obor hodnot je uzavřený interval od nuly do jedné.[11]

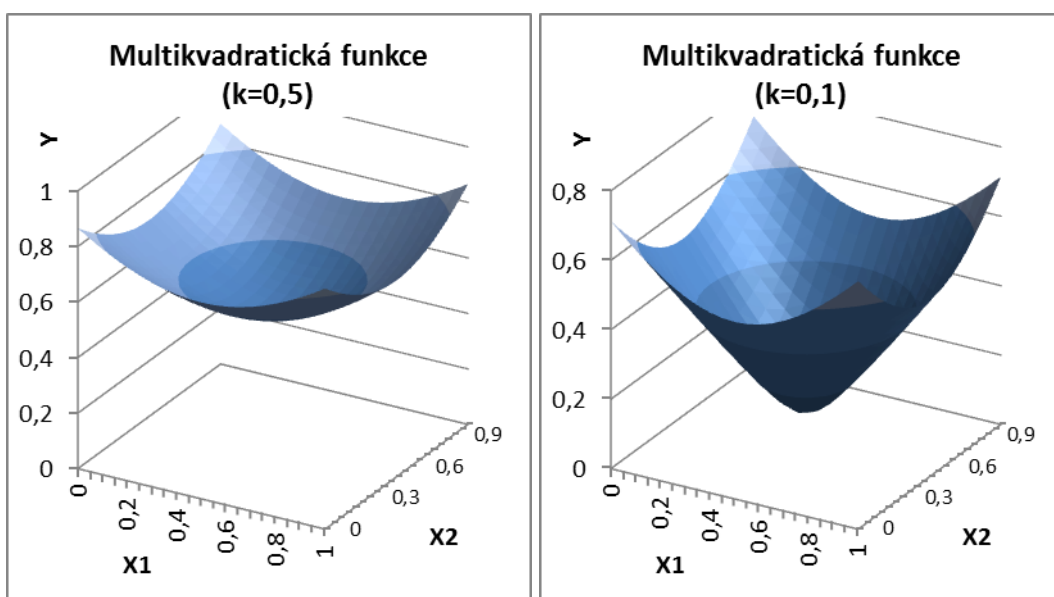


Obr. 2: Průběh Gaussovy funkce pro dvě různá nastavení poloměru

Zdroj: Vlastní zpracování

Multikvadratická funkce je na rozdíl od zbylých dvou funkcí jako jediná rostoucí se vzdáleností od centra. Zdola je omezena nulou a se vzdáleností od centra roste do nekonečna. Parametr k určuje strmost průběhu funkce. Průběh multikvadratické funkce pro dva různé parametry k zobrazuje obr. 3. Tvar funkce je [11]

$$\varphi(r) = (r^2 + k^2)^{1/2} \text{ pro } k > 0, r \in \mathbb{R}. \quad (2.4)$$

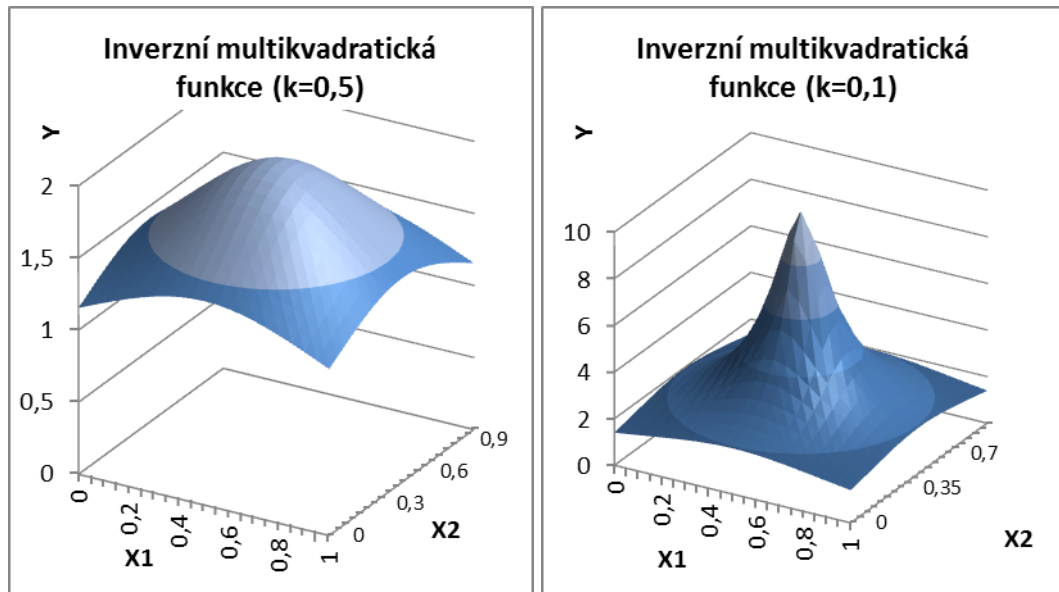


Obr. 3: Průběh multikvadratické funkce pro $k=0,5$ a $0,1$

Zdroj: Vlastní zpracování

Inverzní multikvadratická funkce je definována vztahem (2.5). Hodnota funkce klesá se vzdáleností od středu a je omezena zdola nulou. Parametr k ovlivňuje nejen tvar funkce, ale i maximální hodnotu, které funkce dosahuje ve svém vrcholu, viz obr. 4. Inverzní multikvadratická funkce je definována předpisem [11]

$$\varphi(r) = \frac{1}{(r^2+k^2)^{1/2}} \text{ pro } k > 0, r \in \mathbb{R}. \quad (2.5)$$



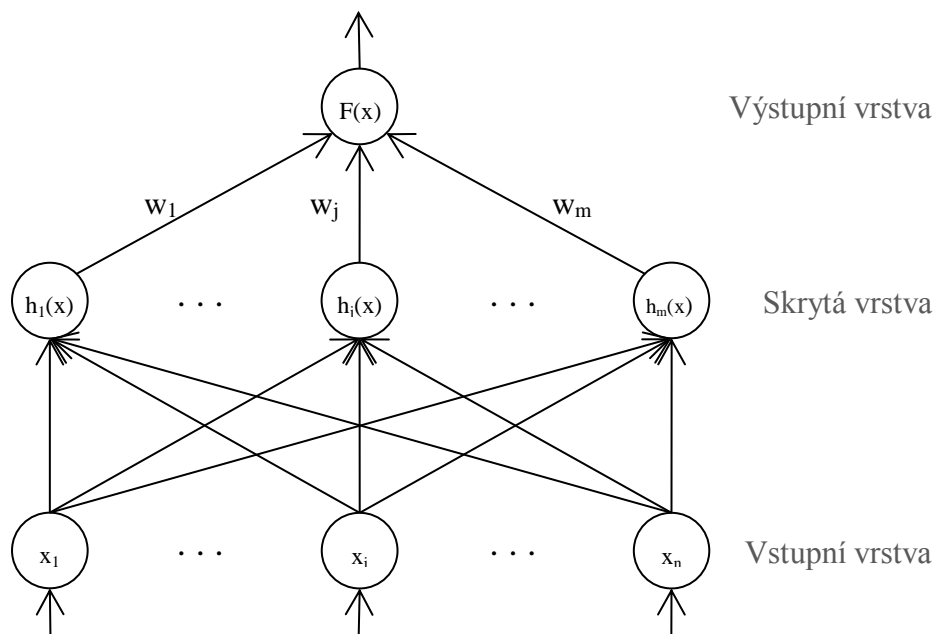
Obr. 4: Průběh inverzní multikvadratické funkce pro $k=0,5$ a $0,1$

Zdroj: Vlastní zpracování

Parametr σ Gaussovy funkce je nazýván poloměr. Multikvadratická ani inverzní multikvadratická funkce nemají poloměr, ale parametr k , jenž ovlivňuje strmost a hraniční hodnoty funkce. Při zakreslení vrstevnic ale tyto parametry ovlivňují míru ohodnocení vzhledem ke vzdálenosti ke středu. Dále v textu je parametr k i σ označen jako střed radiálně bazické funkce.

2.3. Struktura RBF neuronové sítě

RBF neuronové sítě typu využívají mírně odlišnou strukturu oproti MLP. Jako aktivační funkci neuronu využívají radiálně bazické funkce, ve většině případů obsahují jen jednu skrytou vrstvu a učí se s učitelem. Každý neuron je vždy propojen se všemi prvky vrstvy předchozí i následující. S výjimkou poslední vrstvy se zde nevyužívají váhy synapsí. Do všech neuronů v dané vrstvě vždy putuje stejný signál a právě nastavení centra neuronu určuje, jak bude vyhodnocen. [18]



Obr. 5: Struktura neuronové sítě typu RBF

Zdroj: Vypracováno podle [22]

RBF neuronová síť se skládá, ze vstupní vrstvy, alespoň jedné skryté vrstvy a vrstvy výstupní. Dále budeme uvažovat neuronovou síť s jednou skrytou vrstvou tvořenou m neurony, viz obr. 5. Na vstup neuronové sítě jsou přiváděny hodnoty veličiny X . Ty jsou pomocí synapsí distribuovány do všech neuronů skryté vrstvy. Zde jsou transformovány pomocí radiálně bazických funkcí. Každý neuron má nastaven vlastní střed či středy (c_j) a poloměr neuronu (r_j). Výstupem z j -tého neuronu skryté vrstvy je hodnota $h_j(x)$. Synapse mezi skrytou a výstupní vrstvou jsou ohodnoceny pomocí lineárních vah (w_j). Výstup neuronové sítě $F(x)$ lze vypočítat jako lineární kombinaci vah a výstupu ze skryté vrstvy podle vztahu [18]

$$F(\bar{x}) = \sum_{j=1}^m w_j h_j(\bar{x}) = \sum_{j=1}^m w_j \varphi(\|\bar{x} - \bar{c}_j\|). \quad (2.6)$$

2.4. Učení RBF neuronové sítě

Pro neuronovou síť typu RBF, u níž jsou známá centra radiálně bazických funkcí (c_i) a jejich poloměry (r_i) spočívá učení ve stanovení lineárních vah (w_i) mezi skrytou a výstupní vrstvou. Jedním ze způsobů, jak určit parametry c_i a r_i , se zabývá kapitola 3.

Data (X), reprezentující nezávislé proměnné, jsou přivedena na vstup neuronové sítě a lze je popsat maticí o p řádcích a n sloupcích, kde $p, n > 0$. Necht' existuje neuronová síť typu RBF s n vstupy, jedním výstupem a m neurony v jediné skryté vrstvě jako je tomu na obr. 5. Dále existuje vektor vstupu \bar{x}_i (2.7) a jemu odpovídající hodnota výstupu y_i . Závislá proměnná y_i reprezentuje ideální hodnotu, které by měla neuronová síť pro daný vstup x_i dosáhnout. Jedná se tedy o učení s učitelem. Vektor \bar{x}_i je definován

$$\bar{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{in} \end{bmatrix}, i = 1, 2, \dots, p. \quad (2.7)$$

Poté lze definovat vektor [22]

$$\bar{h}_i = \begin{bmatrix} h_1(\bar{x}_i) \\ h_2(\bar{x}_i) \\ \vdots \\ h_m(\bar{x}_i) \end{bmatrix}, i = 1, 2, \dots, p. \quad (2.8)$$

Vektor \bar{h}_i reprezentuje výstupy neuronů skryté vrstvy RBF neuronové sítě, pokud je na vstup přiveden vstupní vektor \bar{x}_i . Vektory h_i (2.8) tvoří jednotlivé řádky matice [11] [22]

$$H = \begin{bmatrix} h_1(\bar{x}_1) & h_2(\bar{x}_1) & \dots & h_m(\bar{x}_1) \\ h_1(\bar{x}_2) & h_2(\bar{x}_2) & \dots & h_m(\bar{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ h_1(\bar{x}_p) & h_2(\bar{x}_p) & \dots & h_m(\bar{x}_p) \end{bmatrix}. \quad (2.9)$$

Stejným způsobem je sestaven i výstupní vektor \bar{y} (2.10), kde hodnota y_i reprezentuje ideální hodnotu pro vstupní hodnoty \bar{x}_i [11]

$$\bar{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}. \quad (2.10)$$

Vektor lineárních vah reprezentuje synaptické váhy mezi skrytou a výstupní vrstvou [11]

$$\bar{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix}. \quad (2.11)$$

Výstup neuronové sítě \bar{y} je lineární kombinací matice H (2.9) a vektoru lineárních vah \bar{w} (2.11). Řešený problém (2.6) lze zapsat rovnicí [11]

$$\begin{bmatrix} h_1(\bar{x}_1) & h_2(\bar{x}_1) & \dots & h_m(\bar{x}_1) \\ h_1(\bar{x}_2) & h_2(\bar{x}_2) & \dots & h_m(\bar{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ h_1(\bar{x}_p) & h_2(\bar{x}_p) & \dots & h_m(\bar{x}_p) \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}. \quad (2.12)$$

Soustavu (2.12) lze přepsat do zjednodušeného tvaru

$$H * \bar{w} = \bar{y}. \quad (2.13)$$

Vzniká soustava p rovnic o m neznámým, kde $m < p$. Znamé jsou hodnoty matice H a vektoru \bar{y} . Neznámé jsou hodnoty lineárních vah w_i . Pro jejich výpočet se využívá metoda nejmenších čtverců [22]

$$\bar{w} = (H^T H)^{-1} H^T \bar{y}. \quad (2.14)$$

Použití metody nejmenších čtverců je velkou výhodou RBF neuronových sítí oproti MLP, pro jejichž optimalizaci je nutné použít iterativní postupy. Soustava (2.14) má řešení jen pokud jsou vektory matice H lineárně nezávislé. V jiném případě není možné vytvořit inverzní matici $(H^T H)^{-1}$ a tedy vypočítat vektor \bar{w} . [22]

Poznámka: Při provádění experimentů s učením neuronové sítě, nebylo u 150 z celkových 156 271 jedinců možno vypočítat inverzní matici. Toto číslo odpovídá asi 0,1%. Jednu generaci tvořilo 20 jedinců. Z každé třetí generace tedy průměrně odpadne 1 jedinec. Snižující populace omezuje možnosti vývoje a přechodu na nové optimum. Proto byla do genetického algoritmu přidána operace doplnění o chybějící jedince.